The international journal of science / 17 September 2020

nature

VIKING VOYAGERS

Genome sequences map ancient maritime travels

Coronavirus
The race to determine the long-term effects of COVID-19

Lessons from history
How scientists used big data to sway voters in the 1960s

Star survivor
Jupiter-sized exoplanet in tight orbit around a white dwarf

Vol. 585, No. 7825
nature.com

The international journal of science / 17 September 2020

nature

A pandemic is no time to cut the ERC's funding

The European Research Council will be crucial to a post-COVID world. Slashing its budget would be a senseless act.

JOHN THYS/AFP/GETTY

European Union member states want to cut funds for Europe's top research agency.

Jean-Pierre Bourguignon is furious. The mathematician is interim president of the European Research Council (ERC), and is outraged by proposals that the agency's budget for 2021–27 is to be cut by €1.3 billion (US$1.5 billion), a reduction of almost 10% from the €14.7 billion that had been agreed by EU leaders in May. "I don't understand it," he told *Nature*. He wants the decision reversed. So do we.

The EU has seen more than 2.5 million cases of the coronavirus, leading to the deaths of more than 142,000 people – out of 925,000 worldwide. At a time like this, you would think that the continent's leaders would want to strengthen the ERC, whose grant recipients are and will be key to understanding SARS-CoV-2, defeating COVID-19 and rebuilding societies and economies during and after the pandemic. But the leaders plan to cut back.

Created in 2007, the ERC is Europe's main funding agency for fundamental research. It is investigator-driven, and the benefits show. Whereas politicians have been slow or late to anticipate and respond to the pandemic, 180 existing ERC projects have been found to be highly relevant to the crisis. ERC investigators are ahead of the curve.

> **"Politicians have been slow to respond to the pandemic. ERC investigators are ahead of the curve."**

Unexpected setback

The council's main difficulty is that its fortunes are tied to those of the EU's larger research and innovation funding programme, Horizon Europe. In previous years, both budgets had been rising. But now the pandemic is devastating economies and, with the United Kingdom no longer in the EU, its contribution will be absent.

In 2018, the European Commission proposed €94.1 billion for Horizon Europe, an increase on the €80-billion budget for the 2014–20 funding programme (known as Horizon 2020). But in July this year, EU leaders chopped that back to €81 billion, including a €5-billion fund for COVID-related research. As a consequence, the ERC's budget will also be cut, even though little of the extra funding is expected to flow to the type of work that the ERC supports, such as developing models to track virus transmission, researching technologies for use in diagnostics and studying human behaviour in a pandemic.

The ERC's other challenge is that returns to society from fundamental research are not always immediately obvious to policymakers – particularly when compared with returns from other parts of the Horizon Europe budget, such as those that support climate science, cancer research and commercial partnerships.

Protecting funding for basic science during a time of budget cuts is monumentally difficult for any research agency, but a turbulent six months for the ERC's leadership has made it harder still.

In April, the agency's then-president, nanoscientist Mauro Ferrari, resigned after three months in office, at just the time when the agency needed to strengthen its coalition of support ahead of budget discussions. Previous president Bourguignon returned in an interim capacity on 27 July – days after the crucial meeting of EU leaders at which budget cuts were proposed.

The ERC is seen as stellar by the standards of basic research agencies. According to the latest evaluation report, almost one-fifth of projects report a breakthrough and more than half lead to a major scientific advance (see go.nature.com/3iyhn9i). Some countries – notably Poland – have even remodelled how they award grants to mirror the ERC's approach.

About 25% of all patents filed by projects supported by Horizon 2020 have come from ERC projects, even though commercialization of research is not the agency's main aim. Bourguignon and his colleagues rightly argue that many advances in fundamental research ultimately contribute to innovation and benefit society. But that is a hard message to get across at a time of constrained funding and competing priorities.

Winds of change

The ERC has also been buffeted by Europe's broader political cross-winds. During previous budget-setting periods, it was able to draw on the support of research and finance ministers from Europe's three biggest economies: Germany, France and the United Kingdom. But the United Kingdom has left the EU; and Germany, for now, is unable to provide its usual strong public backing. Since July, it

nature

has held the rotating presidency of the Council of the European Union, the EU body representing member states' governments. In a statement, Germany's research ministry has said that it supports the ERC but cannot take a position during budget negotiations.

Still, the ERC retains strong support from the European Parliament, from the EU's smaller countries and from research and university leaders. That is why Bourguignon is right to take his case for support directly to these constituencies, which he has been doing. But time is short: the budget will be finalized before the end of this month.

The ERC is a rare success story in multilateral research funding. Its generous starting grants have had a profound impact on the quality of research in Europe. It has helped more experienced scientists to mature as researchers and mentor new talent. That talent is needed to tackle today's crises – and tomorrow's, too.

For their campaign to succeed, the ERC and its supporters need the research community and politicians across Europe to make a stronger case, especially to EU member states' ministries of finance. France and Germany have backed the ERC from the start. Now is not the time to dilute that support for an agency that will be essential to a post-COVID future.

Keep collaboration open when doors are closing

As some countries begin to raise barriers to international collaboration, scientists in the S20 engagement group are right to keep them down.

One by one, doors to international collaboration in research are starting to close.

The US government is leaving the World Health Organization and continuing its crackdown on scientists with connections to China (see page 335). China's government, meanwhile, is ending a policy that actively encouraged researchers to publish with colleagues in other countries.

In the European Union, some leaders have been suggesting that the flagship Horizon Europe research-funding programme should put more conditions on international participation – a dismaying development for an institution founded to strengthen bonds and protect against conflict.

At the beginning of this month, the European Commission published a foresight study aimed, in part, at achieving what it is calling 'technological sovereignty', a phrase that would have been unthinkable even a year ago.

The report finds that the EU has become overly reliant on other countries, especially China, for supplies of crucial raw materials – including graphite, cobalt and lithium – that are needed in batteries and fuel cells, as well as in solar and wind-energy technologies. As fossil-fuel use declines, the EU will need nearly 60 times as much lithium by 2050 as it does today, according to one scenario. It will be looking for ways to bring mining of these materials – and the manufacturing processes they are involved in – closer to home. All of this suggests that the curtain is about to fall on an era of expanding international collaboration in research and technology.

But one group of researchers is sensibly keeping lines of communication open. On 26 September, Saudi Arabia will host the S20 – a meeting of scientists in advance of the G20, the annual gathering of heads of government of the world's 20 biggest economies, due to take place in Riyadh in November.

With science in the spotlight and with research being essential to ending the global coronavirus crisis, the S20 has been conducting a foresight exercise for global benefit. The aim is to assess how all countries could become more resilient to external shocks, such as pandemics, and how they can prepare for the transition to sustainable development. The S20 canvassed expert and lay opinions from around the world, surveyed academic literature and held evidence sessions to discuss what they found.

The final results are due to be published in time for the 26 September meeting, but an interim paper seen by *Nature* makes its timely message clear. The world is now more interconnected than at any time in human history, which means international research collaboration must be central to any ambition to understand how to make societies more resilient.

> **"Societies that seek to erect barriers will find it tougher to withstand sudden shocks."**

It's the right message. Societies that seek to erect barriers – for example, by restricting the flow of ideas – will find it tougher to withstand sudden shocks than will those that are open to sharing what they know, from genome sequences and clinical-trial results to designs for personal protective equipment and source code for contact-tracing apps.

The question is whether the intended audience of politicians and policymakers is ready to listen. Right now, it is hard to see the leaders of the G20 nations pivoting to adopt a more collegial approach to dealing with the pandemic. Too often, it's every country for itself. Take vaccine purchasing as an example. G20 governments, led by the United States, the United Kingdom and the EU, have pre-ordered more than two billion doses. The United Kingdom has purchased 340 million doses – 5 for each citizen – which will leave limited supplies for low- and middle-income countries.

Often, when researchers are involved in providing advice to policymakers – as in the current pandemic – it is deemed necessary for them to step back from decisions based on that advice, on the grounds that research stops where politics and policy begin. But there are exceptions: when countries unilaterally put up barriers to collaboration, researchers cannot remain silent.

That makes the key message of this year's S20 meeting more important than ever: the shifting sands of geopolitics must not affect the relationships that power research.

A personal take on science and society

World view

MARTHA LINCOLN

By Martha Lincoln

A special self-image is no defence against COVID-19

Many countries that see themselves as distinctive have handled the pandemic badly.

As an anthropologist who has studied disease outbreaks in Vietnam, I've been moved by the contrast between the experience of COVID-19 there and in the United States. By late April, my friends in Hanoi were posting pictures of celebrations and joyfully announcing "Social distancing is over!" I'm relieved that infection rates in Vietnam remain low, but their posts seem to come from a parallel universe as I and my family and friends in the United States continue to shelter in place.

Just last year, the United States was considered one of the countries best equipped to confront a virus such as SARS-CoV-2. Others included the United Kingdom, Brazil and Chile – nations ranked by the comprehensive Global Health Security (GHS) Index as being among the world's most prepared. Yet since the pandemic began, these countries have delivered some of the worst outcomes. The United States leads the world in both total cases and total deaths; Brazil's fatalities are second. Chile's per-capita cumulative case rate is the second-highest in Latin America, and the United Kingdom has the highest rate of COVID-19 deaths per capita of all the G7 countries. What might explain these staggering failures?

One thing these countries have in common is 'exceptionalism' – a view of themselves as outliers, in some way distinct from other nations. Their COVID-19 responses suggest that exceptionalist world views can be associated with worse public-health outcomes. Researching this association could help in redefining preparedness and allow more accurate prediction of pandemic successes and failures.

The United Kingdom's decision to leave the European Union is recent evidence that the country – or a large part of it – wants to go it alone. In the early months of the pandemic, Prime Minister Boris Johnson disregarded advice against shaking hands, and the government even considered allowing the virus to spread in pursuit of herd immunity. These actions telegraphed hubris about the country's ability to withstand a public-health crisis.

In the United States, the White House has projected exceptionalist world views in many ways, including by pulling out of the World Health Organization and claiming the virus would disappear "like a miracle". Overconfidence in the nation's ability to respond to COVID-19 is seen at all levels of society, from cuts to pandemic-readiness programmes to people refusing to wear masks in public.

Brazil's populist leader Jair Bolsonaro suggested in March that Brazilians were tough enough to survive infection, so no mandatory precautions were necessary. A chaotic national response allowed the epidemic to flourish. Chilean exceptionalism has been invoked to describe the nation's stable democratic institutions, competent judiciary and thriving free-market economy, but COVID-19 infections surged after reaching low-income communities. Although Chile has a robust health-care system, its epidemiological outcomes reveal troubling levels of inequity. The country's self-flattering image could have caused its leaders to underestimate its vulnerability to the virus.

> **"The pandemic provides a natural experiment on the public-health effects of hubris."**

The pandemic provides a natural experiment on the public-health effects of hubris. One way for researchers to measure and compare exceptionalist world views could be to study public attitudes through surveys and interviews. Exceptionalism could also be identified in what a country's leaders say to the public: do their messages emphasize national specialness, or membership of the international community? Researchers could also examine pandemic responses, assuming that exceptionalist countries will be less likely to learn from other nations. Yet more evidence might come from analysing the media: do news stories describe a country's experience as unique, or draw parallels with experiences elsewhere? Such work could explore whether exceptionalism predicts worse performance in disease control. Instead of relying on untested assumptions about preparedness, as the GHS Index rankings did, researchers could consider actual outcomes.

The analysis would need to look at a variety of possible drivers of pandemic outcomes, to safeguard against cherry-picking. However, it could draw lessons from understudied success stories. Last year's GHS Index rated Vietnam 50th of 195 countries, yet as of 6 September, the country's death toll stood at just 35. An analysis of 36 countries' COVID-19 responses, published last month by the FP Group, a news organization based in Washington DC, ranked Senegal – another lower-middle-income country – second. The United States came 31st.

Vietnam never presumed it would have special protection against disease. Its leaders took no chances in responding to reports of a strange pneumonia in Wuhan, China, and acted decisively to quarantine, test and trace the contacts of early cases. Other nations that exceeded expectations in pandemic response include Cuba and Thailand, which had, as of 2 September, limited deaths to double digits.

In Greek myth, hubris is punished by the goddess Nemesis; in disease control, a hubristic world view risks a particularly vengeful nemesis. Overconfidence in national specialness has led to lack of preparedness, prevented collaboration with global health agencies and limited opportunities to learn from the experience of other countries. By identifying a missing variable in pandemic preparedness – the way nations see themselves – scholars could help to develop a more accurate metric for national readiness to fight disease.

Martha Lincoln is a cultural and medical anthropologist and assistant professor at San Francisco State University, California.
e-mail: mlincoln@sfsu.edu

News in brief

KIDS HIT HARD BY COVID-19 HAVE UNIQUE IMMUNE PROFILE

Most children infected with the new coronavirus show few signs of illness, if any. But a few children are struck by a severe form of COVID-19 that can cause multiple organ failure and even death. Now, scientists have begun to tease out the biology of this rare and devastating condition, called multisystem inflammatory syndrome in children, or MIS-C.

Doctors have diagnosed hundreds of cases of MIS-C, which shares some similarities with the childhood illness Kawasaki's disease. To understand MIS-C's biological profile, Petter Brodin at the Karolinska Institute in Stockholm and his colleagues looked at 13 children with MIS-C, 28 children with Kawasaki's disease and 41 with mild COVID-19 (C. R. Consiglio *et al.* *Cell* https://doi.org/d8fh; 2020). The researchers found that compared with children with Kawasaki's disease, those with MIS-C have lower levels of an immune chemical called IL-17A, which has been implicated in inflammation and autoimmune disorders.

Unlike all the other children studied, children with MIS-C had no antibodies to two coronaviruses that cause the common cold. This deficit might be implicated in the origins of their condition, the authors say.

MORE THAN 100 JOURNALS HAVE VANISHED

Scholarly journals are supposed to provide a lasting record of science. But over the past two decades, 176 open-access journals — and many of the papers published in them — have disappeared from the Internet, according to an analysis published on 27 August (M. Laakso *et al.* Preprint at https://arxiv.org/abs/2008.11933; 2020).

A team led by Mikael Laakso, an information scientist at the Hanken School of Economics in Helsinki, manually collected lists of journals from databases such as the Directory of Open Access Journals and the Keepers Registry, to track down titles that had disappeared between 2000 and 2019 without being enrolled in digital preservation services. Journals were considered "vanished" if less than 50% of their content was still freely available online.

More than half of the 176 vanished journals they identified were in the social sciences and humanities, although life sciences, health sciences and physical sciences and mathematics were also represented. Eighty-eight of the journals were affiliated with a scholarly society or research institution.

The analysis also identified 900 journals that are still online but seem to have stopped publishing papers, so might vanish in the near future.

DISCOVERER OF NEURAL CIRCUITS FOR PARENTING WINS US$3-MILLION PRIZE

Discovering the "on-and-off switch" for good parenting in male and female mouse brains has earned Catherine Dulac (pictured) one of this year's US$3-million Breakthrough prizes — the most lucrative awards in science and mathematics. Dulac, a molecular biologist at Harvard University in Cambridge, Massachusetts, and her team provided the first evidence that male and female mouse brains have the same neural circuitry associated with parenting, which is just triggered differently in each sex. "It went against the dogma that for decades said that male and female brains are organized differently," says biologist Lauren O'Connell at Stanford University, California.

Three other $3-million life-sciences awards were also announced. David Baker at the University of Washington in Seattle won for developing the Rosetta software to design synthetic proteins for therapeutics. Dennis Lo at the Chinese University of Hong Kong in Shatin was recognized for discovering that fetal DNA is present in maternal blood — a finding that led to the development of safer non-invasive prenatal tests for disorders such as Down's syndrome.

And Richard J. Youle at the US National Institute of Neurological Disorders and Stroke in Bethesda, Maryland, won for uncovering the role of two proteins in Parkinson's disease.

The Breakthrough Prize in Mathematics went to Martin Hairer at Imperial College London for his work on stochastic partial differential equations, which describe how complex systems evolve when random influences have to be taken into account.

The Breakthrough Prize in Fundamental Physics went to Eric Adelberger, Jens Gundlach and Blayne Heckel, all at the University of Washington, for their pendulum experiments showing that Isaac Newton's law of gravity still holds, even down to scales of just 52 micrometres. And a special award in fundamental physics recognized the life's work of Steven Weinberg of the University of Texas at Austin, one of the developers of the framework unifying the electromagnetic force with the weak nuclear force.

L TO R: DANIEL MIHAILESCU/AFP/GETTY; RICK FRIEDMAN/CORBIS/GETTY

News in focus

NELSON ALMEIDA/AFP/GETTY

Some 18,000 people globally have received the Oxford AstraZeneca vaccine so far.

RELIEF AS CORONAVIRUS VACCINE TRIALS RESTART — BUT TRANSPARENCY CONCERNS REMAIN

UK trials of the Oxford and AstraZeneca vaccine have resumed after a brief pause, yet key details of the events involved have not been released.

By David Cyranoski & Smriti Mallapaty

The UK trials of a leading coronavirus vaccine that were abruptly halted because of safety concerns have restarted.

The University of Oxford and pharmaceutical company AstraZeneca paused enrolment in the global trials of the vaccine on 6 September, after a person participating in the UK trials experienced an adverse reaction. But on 12 September, the university said an independent committee had found that it was safe to restart.

Scientists say that a pause is not uncommon in large trials, and that a speedy resumption of testing was to be expected. The episode shows that care is being taken with the trials, they say.

"Like anybody else who knows the importance of vaccines, I am very happy that the trial will continue," says Klaus Stöhr, a retired influenza researcher who previously headed the World Health Organization's research and epidemiology division for severe acute respiratory syndrome. But some scientists have criticized the trial sponsors for not releasing more information about the reason for the pause and about their decision-making.

The University of Oxford and AstraZeneca have not yet released details of the adverse reaction that led to the trials' pause and how the decision to resume the UK study was made. Regulators in Brazil announced on 12 September that trials of the vaccine have restarted there, but it is unclear when similar trials in South Africa and the United States might also resume.

Marie-Paule Kieny, a vaccine researcher at INSERM, the French national health-research institute in Paris, says she hopes that research groups working on this or other coronavirus vaccines will share more information about clinical-trial holds in future. The transparency bar should be set much higher than this latest example, says Kieny. "When, ultimately, a vaccine will be made available, public trust will be paramount to ensure public-health impact. And trust needs transparency."

Leading vaccine

The vaccine, AZD1222, is one of the leading candidates being developed to protect against the virus that causes COVID-19, and one of a handful of immunizations in the final stages of clinical testing. The pause in global trials sent a shudder around the world.

Such a quick resumption of the trials was the most likely outcome, says Paul Griffin, an infectious-diseases researcher at the University of Queensland in Brisbane, Australia. In large trials, adverse medical events in volunteers are common, and trial holds are designed to ensure that such events are investigated and volunteers are protected, he says. But, most often, it is later decided that the event was probably not related to participation in the trial and does not pose a safety concern to the rest of the volunteers, says Griffin. That seems to be what has occurred in this case, he says.

It can be difficult to pin down the cause of adverse events, says Jonathan Kimmelman, a bioethicist who studies clinical trials at McGill University in Montreal, Canada. "Often, the best you can do is say that there is a possible link, and then proceed with collecting more data and monitoring outcomes," he says.

The University of Oxford said in a press release on 12 September that the pause, which applied to all trials of the vaccine, was necessary "to allow the review of safety data by an independent safety review committee, and the national regulators".

"The independent review process has concluded and following the recommendations of both the independent safety review committee and the UK regulator, the MHRA [Medicines and Healthcare products Regulatory Agency], the trials will recommence in the UK," the statement reads. The university also said that it cannot disclose medical information about the participant's illness for reasons of confidentiality.

It's appropriate not to disclose information, for patient confidentiality and to ensure valid interpretation of the trial results, says Kristine Macartney, the director of Australia's National Centre for Immunisation Research and Surveillance in Sydney.

Lack of details

But Paul Komesaroff, a physician and bioethicist at Monash University in Melbourne, Australia, questions the university's claim that it could not release information about the adverse event on the basis of confidentiality. It is possible to provide information in a manner that avoids identifying a particular individual, but still provides a summary of the clinical issues that arose, and the conclusions the committee reached about the implications for the study, he says. "It is of concern that they sought to avoid doing so," says Komesaroff.

The University of Oxford and AstraZeneca have not yet responded to requests for comment on this criticism.

Although the university and AstraZeneca have not released information about the adverse event to the public, Pascal Soriot, AstraZeneca's chief executive, reportedly told investors on a telephone call last week that a person in the UK trials had developed symptoms of transverse myelitis, according to health-news website STAT. This condition involves inflammation of the spinal cord, which can be triggered by viruses.

But other scientists say there is a good reason why the company hasn't released more details. If information about the trials is released prematurely, it could present a bias to the clinicians involved in them, says Griffin. The integrity of the trials is on the line, he adds. Griffin expects the pause to have little impact on the UK trials' overall timeline.

But it has not been reported when trials of the vaccine in the United States and South Africa will restart. A spokesperson for AstraZeneca told *Nature* that the company "will be guided by health authorities across the globe as to when other clinical trials of the vaccine can resume".

So far, some 18,000 people globally have received the vaccine. Phase III efficacy trials in the United Kingdom, which began in June, aim to recruit 10,000 people, and a phase III trial in Brazil hopes to recruit 5,000 participants. The US trial, which started in August, is aiming to recruit 30,000 participants. A phase I/II safety and efficacy trial in South Africa wants to recruit 2,000 volunteers.

THE UNDERDOG COVID-19 VACCINES THAT THE WORLD MIGHT NEED

Small developers struggle to get their candidates noticed, but they'll be crucial if front runners stumble.

By Ewen Callaway

When it comes to developing vaccines, Peter Palese is no slouch. A virologist at Icahn School of Medicine at Mount Sinai in New York City, he pioneered genetics techniques that are used to make some of the billions of influenza vaccine doses produced annually, and his team has won millions of dollars to develop a universal flu jab.

Palese is developing a COVID-19 vaccine, too. It consists of a bird virus that has been genetically modified to make a protein found on the surface of SARS-CoV-2. The vaccine fully protects mice from an experimental model of COVID-19, according to a preprint[1] (the research has not yet been peer reviewed). It also grows in chicken eggs, like most flu vaccines, so manufacturing could be ramped up using tried-and-tested technology.

Despite its potential, Palese's vaccine has struggled to gain the attention and funding needed to progress to human trials. "We thought this would be the best thing after sliced bread, and people would break down our doors to get it. That's not the case. We are very disappointed," he says.

As leading drug and biotechnology companies rush their COVID-19 vaccines through clinical trials and eye up fast-track regulatory authorization, dozens of underdog vaccines such as Palese's have stalled, or are advancing along a slower, more conventional path.

Scientists acknowledge that it would be a waste of resources to take every candidate to clinical trials. But they argue that it's essential to have a diverse selection of COVID-19 vaccines in development. Early favourites could fail, confer only partial protection or work poorly in certain age groups; high costs and other barriers might make some of the front runners unsuitable for wide-scale deployment in lower-income countries.

"Everyone is rooting for them to succeed beyond anyone's expectation, but it's prudent to think about what happens if they don't," says Dave O'Connor, a virologist at the University of Wisconsin–Madison. "We need to make sure we have back-up plans — and back-up plans to those back-up plans."

Dozens of candidates

There are more than 320 COVID-19 vaccines in development, according to a tally by the Coalition for Epidemic Preparedness Innovation (CEPI) in Oslo, a fund created to finance

EVA MARIE UZCATEGUI/BLOOMBERG VIA GETTY

Dozens of coronavirus vaccine candidates are in clinical trials.

and coordinate vaccines for outbreaks. Most of these are in the early stages of preclinical development; several dozen are in clinical trials, and only a handful have begun final-phase tests for efficacy. "Everybody and their mother has a vaccine. My dogs have two vaccines," says one scientist working on a leading candidate. Although on the face of it this is good news, it also presents challenges. One is determining which candidates should move forward to costly clinical trials: running even a small study to test safety and dosing is beyond the reach of most academic groups, and smaller teams face an uphill struggle to get their candidates noticed.

In some cases, the breakneck pace of COVID-19 vaccine efforts has created openings for academic groups. One of the leading candidates is being developed by the University of Oxford, UK, and drug company AstraZeneca (see page 331). The vaccine is based on a kind of chimpanzee cold virus, called an adenovirus, that has been used to make experimental vaccines against Ebola, malaria and other diseases, allowing Oxford vaccinologists to quickly adapt the platform to a COVID-19 vaccine. Another technology comprises RNA instructions for a coronavirus protein, and two front-runner vaccines are being developed by firms with expertise in that platform.

But neither technology has yet produced licensed vaccines, and there is no guarantee that the candidates will generate strong immunity against the coronavirus, says Michael Diamond, a viral immunologist at Washington University in St. Louis, Missouri, who is working on two early-stage vaccines. One[2] is based on a weakened livestock virus. The other[3] is based on a chimpanzee adenovirus, like the Oxford–AstraZeneca effort.

Diamond's adenovirus vaccine, unlike any of the leading candidates, is designed to be administered through the nose. A team led by Diamond and Washington University cancer biologist David Curiel found[3] that mice given a single dose of the intranasal vaccine were fully protected from SARS-CoV-2, with almost no sign of virus in their upper or lower airways. Mice that received an injection of the same vaccine were only partially protected, echoing animal data from some leading candidates. This was because the intranasal vaccine summoned potent 'mucosal' immune responses that can block the virus at the site of infection in the upper airways, the team says.

> **"We don't have a billion dollars, but we are moving the programme forward and making sure we don't lose time."**

On the basis of such results, Diamond feels that his team has "a mission" to push its vaccines into human trials, to "see if they're going to be one of the last ones standing – even if they're not the first ones out there". His university has completed a deal to license the intranasal vaccine to a manufacturer, but Diamond hasn't yet found anyone to advance his team's livestock-virus vaccine. Pharmaceutical company Merck is developing its own vaccine based on the same virus, which is also the backbone of the Merck Ebola vaccine that was approved in the United States and the European Union last year. Many companies "just don't have the bandwidth, money, the wherewithal or desire to actually pick up additional platforms", says Diamond. "The challenge has been to find partners."

Many of the vaccines gunning for the first approvals won early funding from CEPI, which has so far spent nearly US$900 million on nine COVID-19 candidates. US government agencies including the Biomedical Advanced Research and Development Authority (BARDA) have spent billions of dollars supporting a handful of candidates as part of Operation Warp Speed. But other funders, with their own priorities, are stepping in to help academics turn their experimental vaccines into products.

Global coverage

With many wealthy countries snapping up early supplies of the leading COVID-19 vaccine candidates, some of these teams have set their sights on developing vaccines for the rest of the world.

Neil King, a biochemist at the University of Washington in Seattle, and his team are readying a nanoparticle vaccine for clinical trials, with support from the Bill & Melinda Gates Foundation in Seattle. The effort, which King is leading with University of Washington structural biologist David Veesler, has produced a vaccine consisting of a self-assembling virus-like particle that is dotted with 60 copies of the receptor-binding domain of the spike protein that SARS-CoV-2 uses to enter human cells. In a preprint, the team reported that tiny doses of the vaccine led to whopping immune responses in mice[4].

The jab could be supplied to low- and middle-income countries, says King. It comprises 'recombinant' proteins made using DNA from multiple sources – which are already used as medical products, including insulin, so there is huge global manufacturing capacity for them. 'Virus-like particle' vaccines that self-assemble from these proteins also have a strong track record: existing vaccines against human papillomavirus, a cause of cervical cancer, and hepatitis B are based on the technology. Clinical trials of the nanoparticle vaccine are set to begin in December. "We don't have a billion dollars from BARDA, but we are moving the programme forward and making sure we don't lose time," says King.

Researchers say that funders need to step in to provide guidance and financial support for COVID-19 vaccines. But as much as underdog developers would like to see their vaccines help bring the pandemic to an end, they are still rooting for their better-funded competitors to succeed. "As a human being, my hope is that none of the candidates fail," says King.

1. Sun, W. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2020.07.30.229120 (2020).
2. Case, J. B. *et al. Cell Host Microbe* **28**, 465–474 (2020).
3. Hassan, A. O. *et al. Cell* https://doi.org/10.1016/j.cell.2020.08.026 (2020).
4. Walls, A. C. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2020.08.11.247395 (2020).

US UNIVERSITY WORKERS FIGHT A RETURN TO CAMPUS AMID COVID-19

Faculty members and other campus staff protest against unsafe conditions as institutions reopen.

By Emma Marris

A wave of activism is sweeping US campuses that have reopened after their summer break amid the COVID-19 crisis. Across the country, university workers are pushing back against requirements that they show up on campus alongside undergraduates. The workers, including faculty members and staff who teach in classrooms and laboratories, along with housekeeping staff who clean dormitories, say they are risking their own health.

One group has filed a lawsuit against the University of North Carolina (UNC) system, which includes 16 institutions across the state, claiming that the system has not provided a safe workplace for its staff. Others have staged protests – including 'die-ins', in which demonstrators have simulated coronavirus deaths – to demand remote classes and more COVID-19 testing. In one case, university faculty members passed a 'no confidence' vote to indicate that their chancellor had neglected their concerns and botched the institution's reopening.

"We are seeing a wave of faculty activism, and it is great for the profession," says Irene Mulvey, president of the American Association of University Professors (AAUP), whose chapters function as unions at some institutions and as advocacy organizations at others. "The COVID-19 crisis and the disastrous decisions being made that literally are putting lives at risk are empowering faculty to take action to fight back."

Protest catalysts

The nascent movement has sprung up as COVID-19 infections on US campuses are growing. An outbreak at UNC-Chapel Hill began after the institution opened its doors in early August. Within nine days, it had shifted to remote instruction, and students had begun moving out of dormitories. By 3 September, 1,075 students and 60 employees at the university had tested positive for the coronavirus. Overall, some 61,000 COVID-19 cases have been reported at US universities since late August, according to *The New York Times*.

Campus workers think that universities have based their reopening decisions on their bleak financial outlooks rather than on safety considerations. Federal and state funding for public universities has declined in recent years in the United States, so universities have become increasingly reliant on tuition and fees to keep running. Administrators fear that students, faced with remote learning, might defer their enrolment until 2021. Going online also means forgoing revenue from dining and housing. For instance, on 27 August, UNC-Chapel Hill's chancellor, Kevin Guskiewicz, said that the university will lose US$55 million dollars in housing and dining during the current semester alone.

For universities in this position, the decision to reopen during the pandemic comes down to one of two choices, says Jay Smith, a history professor at UNC-Chapel Hill and the vice-president of its AAUP chapter. "They either have to risk a public-health disaster or face a certain financial calamity."

"It is scary," says a science professor at Northern Arizona University (NAU) in Flagstaff, who asked to remain anonymous to preserve his relationship with the institution's administration. "It's all about the enrolment money. It has made me feel like we are cogs in a wheel, responsible for keeping the enrolment wheel going." In-person classes at NAU started on 31 August, with some students being taught in person and others dialling in remotely, in an alternating fashion.

Universities say they are doing their best to reopen safely, emphasizing that many students still want in-person classes. Some also underscore that staff can request to teach remotely. "All NAU employees, including faculty and staff, can request accommodations or workplace modifications, including potentially teaching/working remotely," a spokesperson for the institution said in an e-mail. But, unless an NAU employee demonstrates they have a disability that might put them at particular risk of COVID-19, there is no guarantee a request will be granted.

Pushback

Some workers even contend that forcing employees to teach in person and clean packed dormitories is illegal. On 10 August, faculty members, graduate students and staff at several institutions in the UNC system filed a class-action lawsuit in Wake County Superior Court, claiming that their rights to a safe workspace had been violated by the system's reopening plans. Named plaintiffs include Zofia Knorek, an ecology graduate student and research assistant at UNC-Chapel Hill, and housekeeper Jermany Alston, also at UNC-Chapel Hill.

Although state laws limit unions' powers in North Carolina, organizations such as the AAUP and UE Local 150, North Carolina's union for public-service workers – representing staff and graduate students – have joined forces to speak out against reopening plans, including supporting the lawsuit.

TY WRIGHT/BLOOMBERG/GETTY

Graduate students at the Ohio State University protested on the first day of classes.

Faculty members and graduate students across the system are advocating for themselves by protesting and by filing the lawsuit, but they're also advocating for the staff — "dining workers, landscaping workers, and so on — who are among the most vulnerable in our communities", says Smith. Workers such as these are exposed to residential students on campus more often than are any other staff.

The UNC system's board of governors did not respond to repeated requests for comment from *Nature*. The board ultimately dictated the system's reopening plans. It is appointed by the Republican-led North Carolina legislature and has been widely criticized by faculty members for prioritizing revenue over safety, and for not delegating more authority to individual institutions in the system.

A growing movement

Activism is also occurring elsewhere in the UNC system. Faculty members at Appalachian State University in Boone, which is still holding in-person classes, passed a resolution expressing "no confidence" in their chancellor for failing to resist the board of governors' mandate to reopen. The chancellor's office did not respond to a request for comment.

Campus workers are organizing to oppose reopenings elsewhere in the country, too. Georgia College in Milledgeville is open and requiring in-person, teaching despite 686 cases in a campus community of just 8,000. On 28 August, graduate students and staff organized under the United Campus Workers of Georgia staged a die-in — in which participants lay down next to temporary grave stones, spaced apart in a nod to social-distancing measures — to protest against the risk they say reopening poses. "It was kind of a sombre event," says Jessica McQuain, a master's student in English at the university who organized the protest.

Melanie DeVore, a palaeobotanist, is teaching nearly 100 students in person this term at Georgia College. To keep her infection risk low, DeVore got permission to teach outside on a deck. She compares the in-person teaching requirement to the 1979 film *Alien*, in which a spaceship crew discover that although their mission is top priority, they themselves are expendable. And she, like many others, attributes the requirement to teach in person to the university's focus on its finances. "We are backed into a corner because of the business model of the universities," she says.

A spokesperson for Georgia College replied to *Nature*'s interview request with a statement: "Georgia College fully supports the freedoms of speech and expression for our faculty, staff and students." The statement goes on to say that "the health and well-being of our students and campus community will always be our top priority".

The Trump administration has accused China of stealing US intellectual property.

MANDEL NGAN/AFP/GETTY

US CRACKDOWN SPURS FEARS OF CHINESE BRAIN DRAIN

An exodus of foreign-born scientists would be a great loss for US science, say research leaders.

By Andrew Silver

Scientists in the United States are concerned that their government's crackdown on foreign interference at universities is driving away scientists of Chinese descent. Their exodus would be a loss for US innovation, according to extensive interviews *Nature* carried out with scientists and research leaders.

"There are certainly people leaving," says Steven Chu, a Nobel-prizewinning physicist at Stanford University in California, who was secretary of energy under former US president Barack Obama.

The research community has been increasingly feeling the effects of US–China political tensions. US politicians — including President Donald Trump — have accused the Chinese government of using students and researchers to illicitly acquire US knowledge and intellectual property, allegations that the Chinese government has repeatedly denied. Since 2018, US government agencies have unveiled increasingly strict visa restrictions for Chinese nationals, and tighter controls on what research can be shared with China.

US researchers with ties to China who are funded by the National Institutes of Health (NIH) or the National Science Foundation (NSF) have also been investigated for potentially violating funding rules. The NIH said in June that it had investigated 189 researchers who might have violated grant or institutional rules on research integrity. Of these researchers, 93% had ties to China and 82% were Asian. And in the past two months, four researchers from China working in the United States have been charged with visa fraud for allegedly failing to declare links to China's military, a development that marks a new chapter in US–China science relations.

The latest arrests are another example of the US government cracking down on Chinese scholars, says Jessica Chen, an immigration lawyer in Houston, Texas, who has been contacted by researchers for help with immigration issues. Chen says the arrests are part of a pattern of actions that have created a fearful atmosphere and made researchers think about leaving. People cannot focus on their work when they are concerned that they might be investigated or accused of spying, she says. "This creates a truly oppressive environment in which to try to perform research."

Several scientists who spoke to *Nature* say they know of researchers with Chinese backgrounds who have left the United States because they felt nervous or unsafe. Alice Huang, a biologist at the California Institute of Technology in Pasadena and vice-president of the 80-20 Educational Foundation, an advocacy group for Asian American equality, says she knows of around four researchers of Chinese descent who were US citizens and have left the country in the past two years. Some left because they felt they were being targeted by the FBI or NIH, or feared being investigated by them. But she thinks the numbers of researchers leaving the United States are much greater than the cases she's heard about. "We are damaging our own scientific enterprise," says Huang.

Chu knows of a Chinese national who earned a US PhD but has accepted a faculty position in China because of a perceived unfriendly environment in the United States. And he says he's heard from researchers working in science, engineering, technology and mathematics (STEM) who feel unwelcome, or who worry about losing out on jobs or competitive funding because of their country of origin. "I'm trying to convince these people not to go back [to China]," he says. "If it wasn't for immigrant scientists, we would be a second-tier STEM country." However, Chu notes that some researchers are leaving for good opportunities in China.

Researchers of Chinese descent in the United States are also increasingly seeking legal advice because they're concerned they'll be investigated by the government or their institution, says Frank Wu, president of Queen's College, City University of New York, who helps researchers to find suitable lawyers. He says that in the past two years, he's gone from receiving no calls from researchers seeking lawyers to receiving dozens of calls. "They're worried their lives will be ruined for no good reason," he says.

It's difficult to measure whether a significant number of ethnic Chinese scientists have been leaving the United States in response to the government crackdown. Those kinds of data aren't routinely collected, says Brad Farnsworth, vice-president for global engagement at the American Council on Education in Washington DC. But he says that ethnic Chinese researchers in the United States have become even more worried about being under scrutiny since Charles Lieber, a chemist at Harvard University in Cambridge, Massachusetts, was arrested in January for allegedly making false statements about his ties to China. "The level of anxiety has definitely gone up," Farnsworth says.

Concerns about racial profiling

Some scientists and US lawmakers have raised concerns that the government crackdown is verging on racial profiling – the practice of targeting people because of their racial or ethnic background. The concerns sparked a formal investigation by Congress's House of Representatives. In February, representatives Jamie Raskin and Judy Chu, both Democrats, sent letters to the FBI and NIH requesting details of practices that they thought to be suggestive of racial profiling, such as reportedly encouraging universities to scrutinize Chinese Americans or researchers with connections to China. The letter to the FBI also mentions a 2018 study that found that 52% of individuals charged by the US Department of Justice with economic espionage since 2009 have been of Chinese heritage (A. C. Kim *Cardozo Law Rev.* **40**, 749–822; 2018). But those people were more than twice as likely to be acquitted or have charges against them dropped compared with non-Asian defendants.

> "If it wasn't for immigrant scientists, we would be a second-tier STEM country."

Raskin told *Nature* by e-mail that he has received responses from the agencies, and had a briefing with the NIH. "While I get the serious national security implications of Chinese government espionage, none of that justifies dragnet-style ethnic profiling of U.S. citizens who are Chinese-American," he says. "What distinguishes us from authoritarian governments is our Bill of Rights and commitment to the civil liberties and equal rights of all citizens."

The agencies have denied that racial profiling is happening. An FBI spokesperson told *Nature* in a statement that it does not conduct investigations based solely on race, ethnicity or national origin into unlawful activity or threats to national security. "It would not be appropriate for the FBI to ask any university, company, or other entity to profile individuals based on their ethnicity," they wrote. The FBI also stated that it does not comment on engagements with Congress.

When asked to comment on the House investigation and the letter from Raskin and Chu, an NIH spokesperson told *Nature* that it does not comment on continuing investigations. The spokesperson noted that most researchers are honest contributors to the advancement of scientific knowledge. But over the past few years, the agency has been made aware of subversive efforts by foreign entities to coax US scientists to violate the terms and conditions of grant awards for personal gain. When the agency identifies threats, it notifies grant institutions and asks them to investigate, they said.

The Department of Justice does not target researchers for prosecution based on their ethnicity, says Adam Hickey, a deputy assistant attorney general at its national security division. But he agrees that many people prosecuted under the department's 'China Initiative', a programme to counter intellectual-property theft or economic espionage involving China, have been people of Chinese heritage. The initiative has led to several prosecutions of academics – mostly involving tax evasion, grant fraud or making false statements about overseas affiliations.

WHY ARCTIC FIRES ARE BAD NEWS FOR CLIMATE CHANGE

Unprecedented wildfires released record levels of carbon, partly because they burnt peatlands.

By Alexandra Witze

Wildfires blazed along the Arctic Circle this summer, incinerating tundra and blanketing Siberian cities in smoke. By the time the fire season waned at the end of last month, the blazes had emitted a record 244 megatonnes of carbon dioxide – that's 35% more than last year, which also set records. One culprit, scientists say, could be peatlands that are burning as the top of the world melts.

Peatlands are carbon-rich soils that accumulate as waterlogged plants slowly decay, sometimes over thousands of years. They are the most carbon-dense ecosystems on Earth; a typical northern peatland packs in roughly ten times as much carbon as a boreal forest. When peat burns, it releases its ancient carbon to the atmosphere, adding to the heat-trapping gases that cause climate change.

Nearly half the world's peatland-stored carbon lies between 60 and 70 degrees north, along the Arctic Circle. The problem with this is that historically frozen carbon-rich soils are expected to thaw as the planet warms, making them even more vulnerable to wildfires and more likely to release large amounts of carbon. It's a feedback loop: as peatlands release more carbon, global warming increases, which thaws more peat and causes more wildfires (see 'Peatlands burning'). A study published last month shows that northern peatlands could eventually shift from being a net sink for carbon to a net source, further accelerating climate change (G. Hugelius *et al. Proc. Natl Acad. Sci. USA* **117**, 20438–20446; 2020).

The unprecedented Arctic wildfires of 2019 and 2020 show that transformational shifts are already under way, says Thomas Smith, an environmental geographer at the London School of Economics and Political Science. "Alarming is the right term."

Fires in Siberia released record-setting amounts of carbon dioxide this year.

KIRILL KUKHMAR/TASS/GETTY

Zombie fires

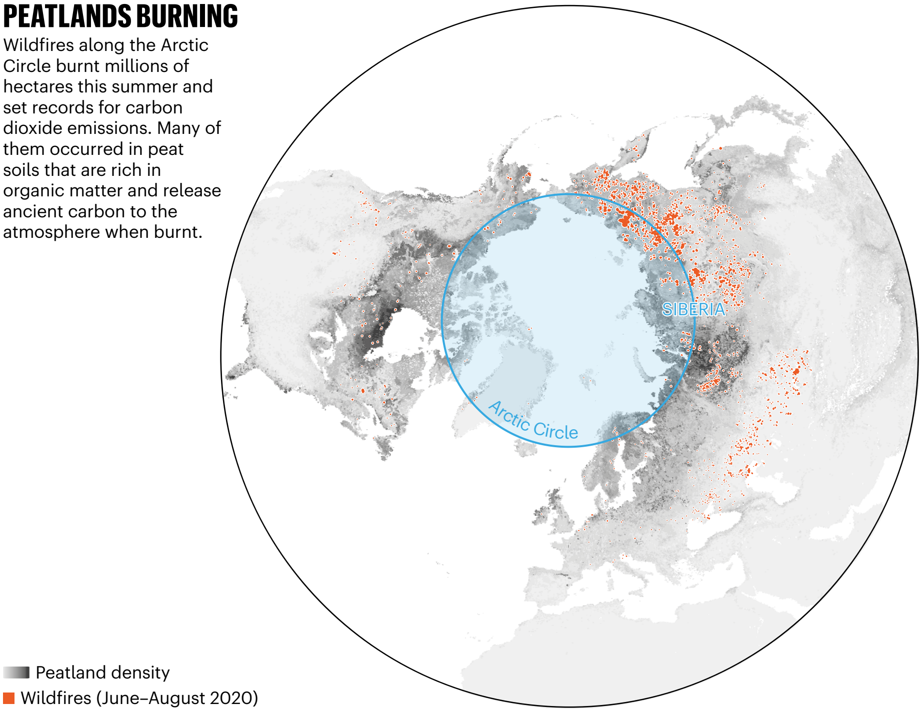
As early as May, there were fires blazing north of the tree line in Siberia, which normally wouldn't happen until around July. One reason is that temperatures in winter and spring were warmer than usual, priming the landscape to burn. It's also possible that peat fires had been smouldering beneath the ice and snow all winter and then emerged, zombie-like, in the spring as the snow melted. Scientists have shown that this kind of low-temperature, flameless combustion can burn in peat and other organic matter, such as coal, for months or even years.

Researchers are now assessing just how bad this Arctic fire season was. The Russian Wildfires Remote Monitoring System catalogued 18,591 separate fires in Russia's two easternmost districts, with a total of nearly 14 million hectares burnt, says Evgeny Shvetsov, a fire specialist at the Sukachev Institute of Forest, which is part of the Russian Academy of Sciences in Krasnoyarsk. Most of the burning happened in permafrost zones, where the ground is normally frozen year-round.

To estimate the record carbon dioxide emissions, scientists with the European Commission's Copernicus Atmosphere Monitoring Service used satellites to study the wildfires' locations and intensity, and then calculated how much fuel each had probably burnt (see go.nature.com/2zk8wcn). Yet even that is likely to be an underestimate, says Mark Parrington, an atmospheric scientist at the European Centre for Medium-Range Weather Forecasts in Reading, UK, who was involved in the analysis. Fires that burn in peatland can be too low-intensity for satellite sensors to capture.

PEATLANDS BURNING

Wildfires along the Arctic Circle burnt millions of hectares this summer and set records for carbon dioxide emissions. Many of them occurred in peat soils that are rich in organic matter and release ancient carbon to the atmosphere when burnt.



SOURCES: COPERNICUS ATMOSPHERE MONITORING SERVICE/EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS; G. HUGELIUS *ET AL. PROC. NATL. ACAD. SCI. USA* **117**, 20438–20446 (2020)

The problem with peat

How much this year's Arctic fires will affect global climate over the long term depends on what they burnt. That's because peatlands, unlike boreal forest, do not regrow quickly after a fire, so the carbon released is permanently lost to the atmosphere.

Smith has calculated that about half of the Arctic wildfires in May and June were on peatlands – and that in many cases, the fires went on for days, suggesting that they were fuelled by thick layers of peat or other soil rich in organic matter (see go.nature.com/3ip4d3y).

And the August study found that there are nearly four million square kilometres of peatlands in northern latitudes. More of that than previously thought is frozen and shallow – and therefore vulnerable to thawing and drying out, says Gustaf Hugelius, a permafrost scientist at Stockholm University who led the investigation. He and his colleagues also found that although peatlands have been helping to cool the climate for thousands of years, by storing carbon as they accumulate, they will probably become a net source of carbon being released into the atmosphere – which could happen by the end of the century.

Fire risk in Siberia is predicted to increase as the climate warms (B. G. Sherstyukov and A. B. Sherstyukov *Russ. Meteorol. Hydrol.* **39**, 292–301; 2014), but by many measures, the shift has already arrived, says Amber Soja, an environmental scientist who studies Arctic fires at the US National Institute of Aerospace in Hampton, Virginia. "What you would expect is already happening," she says. "And in some cases faster than we would have expected."

MARCO DI LAURO/GETTY

A person who has recovered from COVID-19 takes part in a rehabilitation programme in Genoa, Italy.

COVID-19'S LASTING MISERY

Months after infection with SARS-CoV-2, some people are still battling fatigue, lung damage and an array of other symptoms. **By Michael Marshall**

The lung scans were the first sign of trouble. In the early weeks of the coronavirus pandemic, clinical radiologist Ali Gholamrezanezhad began to notice that some people who had cleared their COVID-19 infection still had distinct signs of damage. "Unfortunately, sometimes the scar never goes away," he says.

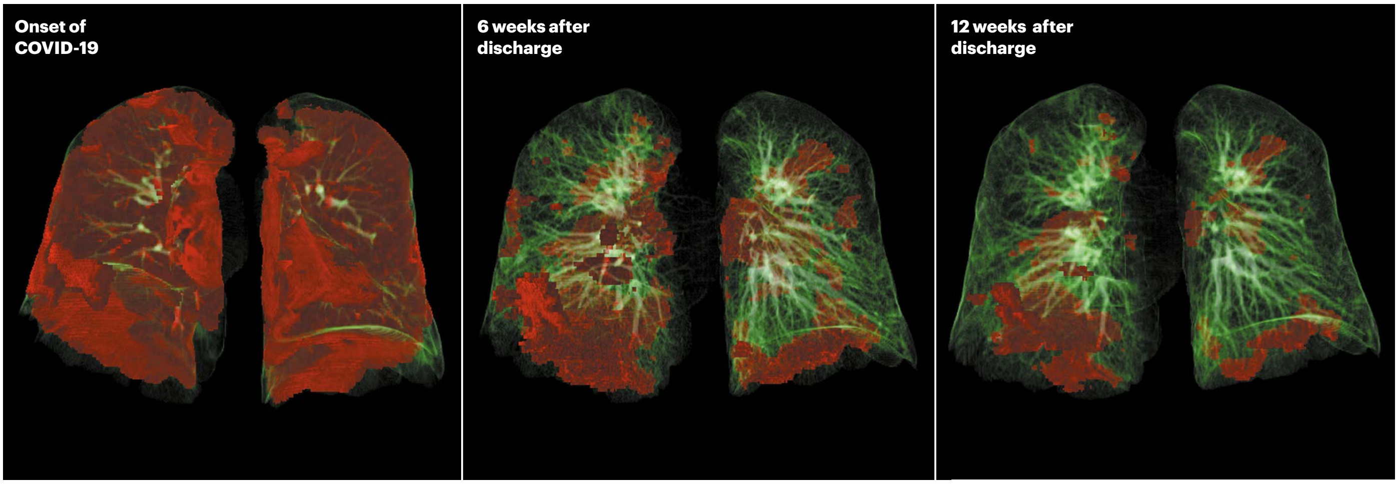
Gholamrezanezhad, at the University of Southern California in Los Angeles, and his team started tracking patients in January using computed tomography (CT) scanning to study their lungs. They followed up on 33 of them more than a month later, and their as-yet-unpublished data suggest that more than one-third had tissue death that has led to visible scars. The team plans to follow the group for several years.

These patients are likely to represent the worst-case scenario. Because most infected people do not end up in hospital, Gholamrezanezhad says the overall rate of such intermediate-term lung damage is likely to be much lower — his best guess is that it is less than 10%. Nevertheless, given that 28.2 million people are known to have been infected so far, and that the lungs are just one of the places that clinicians have detected damage, even that low percentage implies that hundreds of thousands of people are experiencing lasting health consequences.

Doctors are now concerned that the pandemic will lead to a significant surge of people battling lasting illnesses and disabilities. Because the disease is so new, no one knows yet what the long-term impacts will be. Some of the damage is likely to be a side effect of intensive treatments such as intubation, whereas other lingering problems could be caused by the virus itself. But preliminary studies and existing research into other coronaviruses suggest that the virus can injure multiple organs and cause some surprising symptoms.

People with more severe infections might experience long-term damage not just in their lungs, but in their heart, immune system, brain and elsewhere. Evidence from previous coronavirus outbreaks, especially the severe acute respiratory syndrome (SARS) epidemic, suggests that these effects can last for years.

And although in some cases the most severe infections also cause the worst



G. WIDMANN, C. SCHWABL & A. LUGER/INNSBRUCK MEDICAL UNIV.

Lung scans from a 50-year-old show that damage from COVID-19 (red) can improve with time — but many patients have lasting symptoms.

long-term impacts, even mild cases can have life-changing effects — notably a lingering malaise similar to chronic fatigue syndrome.

Many researchers are now launching follow-up studies of people who had been infected with SARS-CoV-2, the virus that causes COVID-19. Several of these focus on damage to specific organs or systems; others plan to track a range of effects. In the United Kingdom, the Post-Hospitalisation COVID-19 Study (PHOSP-COVID) aims to follow 10,000 patients for a year, analysing clinical factors such as blood tests and scans, and collecting data on biomarkers. A similar study of hundreds of people over 2 years launched in the United States at the end of July.

What they find will be crucial in treating those with lasting symptoms and trying to prevent new infections from lingering. "We need clinical guidelines on what this care of survivors of COVID-19 should look like," says Nahid Bhadelia, an infectious-diseases clinician at Boston University School of Medicine in Massachusetts, who is setting up a clinic to support people with COVID-19. "That can't evolve until we quantify the problem."

Enduring effects

In the first few months of the pandemic, as governments scrambled to stem the spread by implementing lockdowns and hospitals struggled to cope with the tide of cases, most research focused on treating or preventing infection.

Doctors were well aware that viral infections could lead to chronic illness, but exploring that was not a priority. "At the beginning, everything was acute, and now we're recognizing that there may be more problems," says Helen Su, an immunologist at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland. "There is a definite need for long-term studies."

The obvious place to check for long-term harm is in the lungs, because COVID-19 begins as a respiratory infection. Few peer-reviewed studies exploring lasting lung damage have been published. Gholamrezanezhad's team analysed lung CT images of 919 patients from published studies[1], and found that the lower lobes of the lungs are the most frequently damaged. The scans were riddled with opaque patches that indicate inflammation, that might make it difficult to breathe during sustained exercise. Visible damage normally reduced after two weeks[1]. An Austrian study also found that lung damage lessened with time: 88% of participants had visible damage 6 weeks after being discharged from hospital, but by 12 weeks, this number had fallen to 56% (see go.nature.com/3hiiopi).

Symptoms might take a long time to fade; a study[2] posted on the preprint server medRxiv in August followed up on people who had been hospitalized, and found that even a month after being discharged, more than 70% were reporting shortness of breath and 13.5% were still using oxygen at home.

Evidence from people infected with other coronaviruses suggests that the damage will linger for some. A study[3] published in February recorded long-term lung harm from SARS, which is caused by SARS-CoV-1. Between 2003 and 2018, Peixun Zhang at Peking University People's Hospital in Beijing and his colleagues tracked the health of 71 people who had been hospitalized with SARS. Even after 15 years, 4.6% still had visible lesions on their lungs, and 38% had reduced diffusion capacity, meaning that their lungs were poor at transferring oxygen into the blood and removing carbon dioxide from it.

COVID-19 often strikes the lungs first, but it is not simply a respiratory disease, and in many people, the lungs are not the worst-affected organ. In part, that's because cells in many different locations harbour the ACE2 receptor that is the virus's major target, but also because the infection can harm the immune system, which pervades the whole body.

Some people who have recovered from COVID-19 could be left with a weakened immune system. Many other viruses are thought to do this. "For a long time, it's been suggested that people who have been infected with measles are immunosuppressed in an extended period and are vulnerable to other infections," says Daniel Chertow, who studies emerging pathogens at the National Institutes of Health Clinical Center in Bethesda, Maryland. "I'm not saying that would be the case for COVID, I'm just saying there's a lot we don't know." SARS, for instance, is known to decrease immune-system activity by reducing the production of signalling molecules called interferons[4].

The virus can also have the opposite effect, causing parts of the immune system to become overactive and trigger harmful inflammation throughout the body. This is well documented in the acute phase of the illness, and is implicated in some of the short-term impacts. For instance, it might explain why a small number of children with COVID-19 develop widespread inflammation and organ problems.

This immune over-reaction can also happen in adults with severe COVID-19, and researchers want to know more about the knock-on effects after the virus has run its course. "It seems there's a lag there for it to get hold of the person and then cause this severe inflammation," says Adrienne Randolph, a senior associate in critical-care medicine at Boston Children's Hospital. "But then the thing is that, long term, when they recover, how long does it take the immune system to settle back to normality?"

Heart of the matter

An over-reactive immune system can lead to inflammation, and one particularly susceptible organ is the heart. During the acute phase of COVID-19, about one-third of patients show cardiovascular symptoms, says Mao Chen, a cardiologist at Sichuan University in Chengdu, China. "It's absolutely one of the short-term consequences."

One such symptom is cardiomyopathy, in which the muscles of the heart become stretched, stiff or thickened, affecting the heart's ability to pump blood. Some patients also have pulmonary thrombosis, in which a clot blocks a blood vessel in the lungs. The virus can also injure the wider circulatory system, for instance, by infecting the cells lining blood vessels[5].

"My major concern is also the long-term impact," says Chen. In some patients, he says, the risk to the cardiovascular system "lingers for a long time". Chen and his colleagues reviewed data from before the pandemic for a study[6] published in May, noting that people who have had pneumonia are at increased risk of cardiovascular disease 10 years later – although the absolute risk is still small. Chen speculates that an over-reactive immune system, and the resulting inflammation, might be involved. However, there is little information on long-term cardiovascular harms from SARS or the related disease Middle Eastern respiratory syndrome (MERS), let alone from SARS-CoV-2.

Studies are now starting. At the beginning of June, the British Heart Foundation in London announced six research programmes, one of which will follow hospitalized patients for six months, tracking damage to their hearts and other organs. Data-sharing initiatives such as the CAPACITY registry, launched in March, are compiling reports from dozens of European hospitals about people with COVID-19 who have cardiovascular complications.

Similar long-term studies are needed to understand the neurological and psychological consequences of COVID-19. Many people who become severely ill experience neurological complications such as delirium, and there is evidence that cognitive difficulties, including confusion and memory loss, persist for some time after the acute symptoms have cleared (see page 342). But it is not clear whether this is because the virus can infect the brain, or whether the symptoms are a secondary consequence – perhaps of inflammation.

Chronic fatigue

One of the most insidious long-term effects of COVID-19 is its least understood: severe fatigue. Over the past nine months, an increasing number of people have reported crippling exhaustion and malaise after having the virus. Support groups on sites such as Facebook host thousands of members, who sometimes call themselves "long-haulers". They struggle to get out of bed, or to work for more than a few minutes or hours at a time. One study[7] of 143 people with COVID-19 discharged from a hospital in Rome found that 53% had reported fatigue and 43% had shortness of breath an average of 2 months after their symptoms started. A study of patients in China showed that 25% had abnormal lung function after 3 months, and that 16% were still fatigued[8].

Paul Garner, a infectious-disease researcher at the Liverpool School of Tropical Medicine, UK, has experienced this at first hand. His initial symptoms were mild, but he has since experienced "a roller coaster of ill health, extreme emotions and utter exhaustion". His mind became "foggy" and new symptoms cropped up almost every day, ranging from breathlessness to arthritis in his hands.

These symptoms resemble chronic fatigue syndrome, also known as myalgic encephalomyelitis (ME). The medical profession has struggled for decades to define the disease – leading to a breakdown of trust with some patients. There are no known biomarkers, so it can only be diagnosed based on symptoms. Because the cause is not fully understood, it is unclear how to develop a treatment. Dismissive attitudes from doctors persist, according to some patients.

People reporting chronic fatigue after having COVID-19 describe similar difficulties. In the forums, many long-haulers say they have received little or no support from doctors – perhaps because many of them showed only mild symptoms, or none at all, and were never hospitalized or in danger of dying.

> **"We don't stick around past the acute stage. We don't look at the long tail of recovery."**

The only way to find out whether SARS-CoV-2 is behind these symptoms is to compare people known to have had the virus with those who have not, says Chertow, to see how often fatigue manifests and in what form. Otherwise there is a risk of lumping together people whose fatigue has manifested for different reasons, and who might need distinct treatments.

Chertow says he is not aware of such a study for COVID-19, but they have been done for other diseases. Following the Ebola epidemic in West Africa in 2014–16, US researchers collaborated with the Ministry of Health in Liberia to perform a long-term follow-up study[9] called Prevail III. The study identified six long-term impacts from Ebola, ranging from joint pain to memory loss. Bhadelia, who treated hundreds of people with Ebola during the outbreak, says that these post-viral symptoms had not previously been recognized. Usually, she says, "we don't stick around past the acute stage. We don't look at the long tail of recovery. It's important to do that, because it tells you more about the virus and its pathophysiology."

The situation is clearer for people who have been severely ill with COVID-19, especially those who ended up on ventilators, says Chertow. In the worst cases, patients experience injury to muscles or the nerves that supply them, and often face "a really long-fought battle on the order of months or up to years" to regain their previous health and fitness, he says. He and his colleagues are now recruiting people with COVID-19 from across the severity spectrum for a long-term follow-up study, assessing their brains, lungs, hearts, kidneys and inflammation responses while they are acutely ill, then during recovery a few weeks later, and again after 6–12 months (see go.nature.com/3mfqqxc).

Once again, there is evidence from SARS that coronavirus infection can cause long-term fatigue. In 2011, Harvey Moldofsky and John Patcai at the University of Toronto in Canada described 22 people with SARS, all of whom remained unable to work 13–36 months after infection[10]. Compared with matched controls, they had persistent fatigue, muscle pain, depression and disrupted sleep. Another study[11], published in 2009, tracked people with SARS for 4 years and found that 40% had chronic fatigue. Many were unemployed and had experienced social stigmatization.

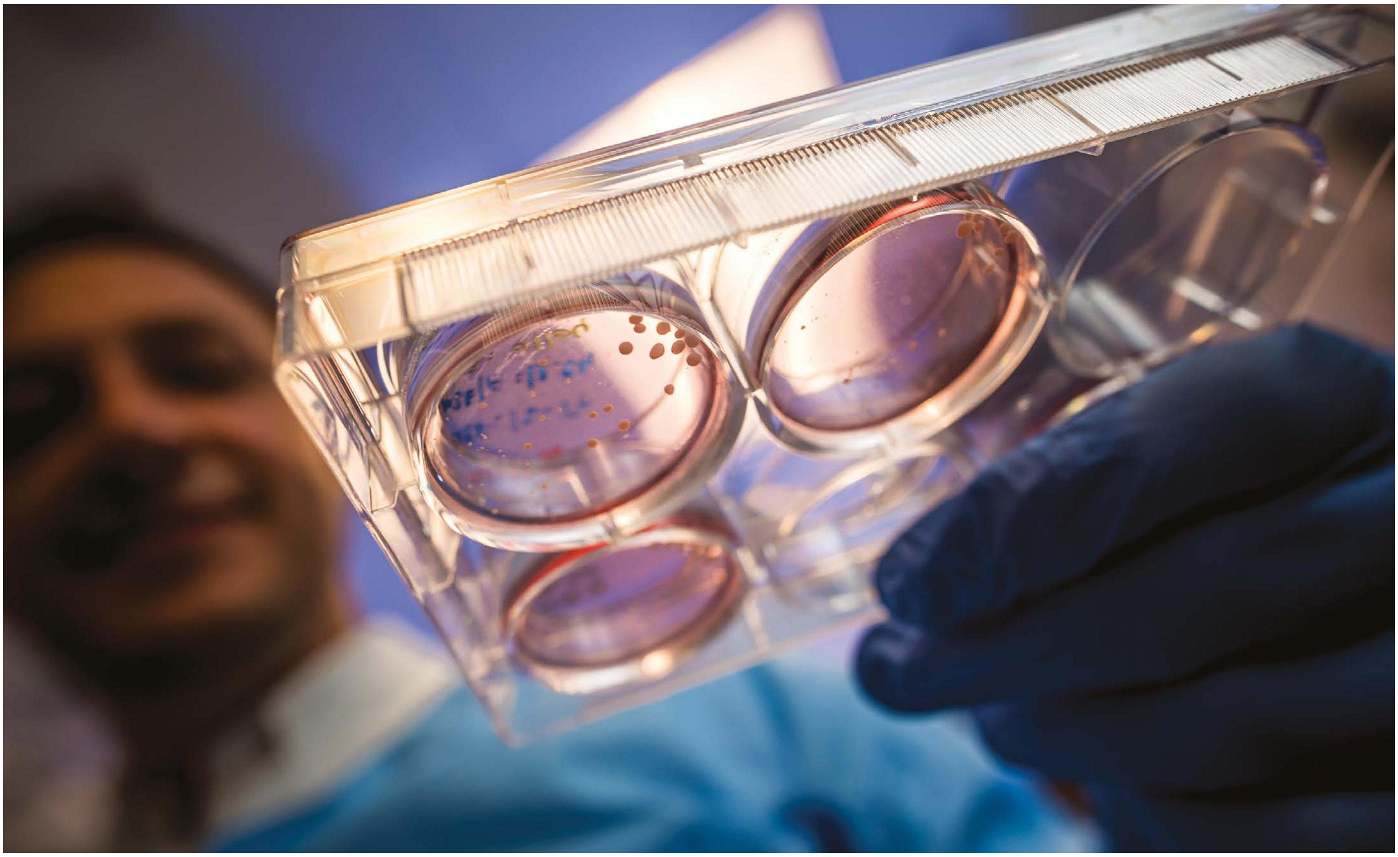
It is not clear how viruses might do this damage, but a 2017 review[12] of the literature on chronic fatigue syndrome found that many patients have persistent low-level inflammation, possibly triggered by infection.

If COVID-19 is such a trigger, a wave of psychological effects "may be imminent", write a group of researchers led by Declan Lyons, a psychiatrist at St Patrick's Mental Health Services in Dublin[13]. In many countries, the pandemic shows no sign of waning, and health systems are already at capacity responding to acute cases. Nevertheless, researchers say it is crucial to start digging into the long-term effects now.

But the answers will not come quickly. "The problem is," says Gholamrezanezhad, "to assess long-term consequences, the only thing you need is time."

Michael Marshall is a science writer based in Devon, UK.

1. Salehi, S., Abedi, A., Balakrishnan. S. & Gholamrezanezhad, A. *Am. J. Roentgenol.* **215**, 87–93 (2020).
2. Weerahandi, H. *et al.* Preprint at medRxiv https://doi.org/10.1101/2020.08.11.20172742 (2020).
3. Zhang, P. *et al. Bone Res.* **8**, 8 (2020).
4. Hu, Y. *et al. J. Virol.* **91**, e02143-16 (2017).
5. Varga, Z. *et al. Lancet* **395**, 1417–1418 (2020).
6. Xiong, T.-Y., Redwood, S., Prendergast, B. & Chen, M. *Eur. Heart J.* **41**, 1798–1800 (2020).
7. Carfì, A. *et al. J. Am. Med. Assoc.* **324**, 603–605 (2020).
8. Zhao, Y.-M. *et al. EClinicalMedicine* **25**, 100463 (2020).
9. The Prevail III study group. *N. Engl. J. Med.* **380**, 924–934 (2019).
10. Moldofsky, H. & Patcai, J. *BMC Neurol.* **11**, 37 (2011).
11. Lam, M. H.-B. *et al. Arch. Intern. Med.* **169**, 2142–2147 (2009).
12. Gerwyn, M. & Maes, M. *Curr. Rheumatol. Rep.* **19**, 1 (2017).
13. Lyons, D. *et al. Ir. J. Psychol. Med.* https://doi.org/10.1017/ipm.2020.40 (2020).



ERIK JEPSEN/UC SAN DIEGO

Some evidence that SARS-CoV-2 can infect the brain comes from ‘organoids’ — clumps of neurons created in a dish.

HOW COVID-19 CAN DAMAGE THE BRAIN

Some people who become ill with the coronavirus develop neurological symptoms. Yet scientists are struggling to nail down the disease’s mental toll.

By Michael Marshall

The woman had seen lions and monkeys in her house. She was becoming disoriented and aggressive towards others, and was convinced that her husband was an impostor. She was in her mid-50s – decades older than the age at which psychosis typically develops – and had no psychiatric history. What she did have, however, was COVID-19. Hers was one of the first known cases of someone developing psychosis after contracting the disease[1].

In the early months of the COVID-19 pandemic, doctors struggled to keep patients breathing, and focused mainly on treating damage to the lungs and circulatory system. But even then, evidence for neurological effects was accumulating. Some people hospitalized with COVID-19 were experiencing delirium: they were confused, disorientated and agitated[2]. In April, a group in Japan published[3] the first report of someone with COVID-19 who had swelling and inflammation in brain tissues. Another report[4] described a patient with deterioration of myelin, a fatty coating that protects neurons and is irreversibly damaged in neurodegenerative diseases such as multiple sclerosis.

“The neurological symptoms are only becoming more and more scary,” says Alysson Muotri, a neuroscientist at the University of California, San Diego, in La Jolla.

The list now includes stroke, brain haemorrhage and memory loss. It is not unheard of for serious diseases to cause such effects, but the scale of the COVID-19 pandemic means that thousands or even tens of thousands of people could already have these symptoms, and some might be facing lifelong problems as a result.

Yet researchers are struggling to answer key questions – including basic ones, such as how many people have these conditions, and who is at risk. Most importantly, they want to know why these particular symptoms are showing up.

Although viruses can invade and infect the brain, it is not clear whether SARS-CoV-2 does so to a significant extent. The neurological symptoms might instead be a result of overstimulation of the immune system. It is crucial to find out, because these two scenarios require entirely different treatments. “That’s why the disease mechanisms are so important,” says Benedict Michael, a neurologist at the University of Liverpool, UK.

Affected brains

As the pandemic ramped up, Michael and his colleagues were among many scientists who began compiling case reports of neurological complications linked to COVID-19.

In a June paper[5], he and his team analysed clinical details for 125 people in the United Kingdom with COVID-19 who had neurological or psychiatric effects. Of these, 62% had experienced damage to the brain's blood supply, such as strokes and haemorrhages, and 31% had altered mental states, such as confusion or prolonged unconsciousness – sometimes accompanied by encephalitis, the swelling of brain tissue. Ten people who had altered mental states developed psychosis.

Not all people with neurological symptoms have been seriously ill in intensive-care units, either. "We've seen this group of younger people without conventional risk factors who are having strokes, and patients having acute changes in mental status that are not otherwise explained," says Michael.

A similar study[1] published in July compiled detailed case reports of 43 people with neurological complications from COVID-19. Some patterns are becoming clear, says Michael Zandi, a neurologist at University College London and a lead author on the study. The most common neurological effects are stroke and encephalitis. The latter can escalate to a severe form called acute disseminated encephalomyelitis, in which both the brain and spinal cord become inflamed and neurons lose their myelin coatings – leading to symptoms resembling those of multiple sclerosis. Some of the worst-affected patients had only mild respiratory symptoms. "This was the brain being hit as their main disease," says Zandi.

Less common complications include peripheral nerve damage, typical of Guillain–Barré syndrome, and what Zandi calls "a hodgepodge of things", such as anxiety and post-traumatic stress disorder. Similar symptoms have been seen in outbreaks of severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), also caused by coronaviruses. But fewer people were infected in those outbreaks, so less data are available.

How many people?

Clinicians don't know how common these neurological effects are. Another study[6] published in July estimated their prevalence using data from other coronaviruses. Symptoms affecting the central nervous system occurred in at least 0.04% of people with SARS and in 0.2% of those with MERS. Given that there are now 28.2 million confirmed cases of COVID-19 worldwide, this could imply that between 10,000 and 50,000 people have experienced neurological complications.

But a major problem in quantifying cases is that clinical studies have typically focused on people with COVID-19 who were hospitalized, often those who required intensive care. The prevalence of neurological symptoms in this group could be "more than 50%", says neurobiologist Fernanda De Felice at the Federal University of Rio de Janeiro in Brazil. But there is much less information about those who had mild illness or no respiratory symptoms.

That scarcity of data means it is difficult to work out why some people have neurological symptoms and others do not. It is also unclear whether the effects will linger: COVID-19 can have other health impacts that last for months, and different coronaviruses have left some people with symptoms for years (see page 339).

Infection or inflammation?

The most pressing question for many neuroscientists, however, is why the brain is affected at all. Although the pattern of disorders is fairly consistent, the underlying mechanisms are not yet clear, says De Felice.

Finding an answer will help clinicians to choose the right treatments. "If this is direct viral infection of the central nervous system, these are the patients we should be targeting for remdesivir or another antiviral," says Michael. "Whereas if the virus is not in the central nervous system, maybe the virus is clear of the body, then we need to treat with anti-inflammatory therapies."

Getting it wrong would be harmful. "It's pointless giving the antivirals to someone if the virus is gone, and it's risky giving anti-inflammatories to someone who's got a virus in their brain," says Michael.

> "The neurological symptoms are only becoming more and more scary."

There is clear evidence that SARS-CoV-2 can infect neurons. Muotri's team specializes in building 'organoids' – miniaturized clumps of brain tissue, made by coaxing human pluripotent stem cells to differentiate into neurons.

In a May preprint[7], the team showed that SARS-CoV-2 could infect neurons in these organoids, killing some and reducing the formation of synapses between them. Work by immunologist Akiko Iwasaki and her colleagues at Yale University School of Medicine in New Haven, Connecticut, seems to confirm this using human organoids, mouse brains and some post-mortem examinations, according to a preprint published on 8 September[8]. But questions remain over how the virus might reach people's brains.

Because loss of smell is a common symptom, neurologists wondered whether the olfactory nerve might provide a route of entry. "Everyone was concerned that this was a possibility," says Michael. But the evidence points against it.

A team led by Mary Fowkes, a pathologist at the Icahn School of Medicine at Mount Sinai in New York City, posted a preprint in late May[9] describing post mortems in 67 people who had died of COVID-19. "We have seen the virus in the brain itself," says Fowkes: electron microscopes revealed its presence. But virus levels were low and were not consistently detectable. Furthermore, if the virus was invading through the olfactory nerve, the associated brain region should be the first to be affected. "We're simply not seeing the virus involved in the olfactory bulb," says Fowkes. Rather, she says, infections in the brain are small and tend to cluster around blood vessels.

Michael agrees that the virus is hard to find in the brain, compared with other organs. Tests using the polymerase chain reaction (PCR) often do not detect it there, despite their high sensitivity, and several studies have failed to find any virus particles in the cerebrospinal fluid that surrounds the brain and spinal cord (see, for example, ref. 10). One reason might be that the ACE2 receptor, a protein on human cells that the virus uses to gain entry, is not expressed much in brain cells[11].

"It seems to be incredibly rare that you get viral central nervous system infection," Michael says. That means many of the problems clinicians are seeing are probably a result of the body's immune system fighting the virus.

Still, this might not be true in all cases, which means that researchers will need to identify biomarkers that can reliably distinguish between a viral brain infection and immune activity. That, for now, means more clinical research, post mortems and physiological studies.

De Felice says that she and her colleagues are planning to follow patients who have recovered after intensive care, and create a biobank of samples including cerebrospinal fluid. Zandi says that similar studies are beginning at University College London. Researchers will no doubt be sorting through such samples for years. Although the questions they're addressing have come up during nearly every disease outbreak, COVID-19 presents new challenges and opportunities, says Michael. "What we haven't had since 1918 is a pandemic on this scale."

Michael Marshall is a science writer based in Devon, UK.

1. Paterson, R. W. *et al. Brain* https://doi.org/10.1093/brain/awaa240 (2020).
2. Kotfis, K. *et al. Crit. Care* **24**, 176 (2020).
3. Moriguchi, T. *et al. Int. J. Infect. Dis.* **94**, 55–58 (2020).
4. Zanin, L. *et al. Acta Neurochir.* **162**, 1491–1494 (2020).
5. Varatharaj, A. *et al. Lancet Psychiatry* https://doi.org/10.1016/S2215-0366(20)30287-X (2020).
6. Ellul, M. A. *et al. Lancet Neurol.* **19**, 767–783 (2020).
7. Mesci, P. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2020.05.30.125856 (2020).
8. Song, E. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2020.06.25.169946 (2020).
9. Bryce, C. *et al.* Preprint at medRxiv https://doi.org/10.1101/2020.05.18.20099960 (2020).
10. Al Saiegh, F. *et al. J. Neurol. Neurosurg. Psychiatry* **91**, 846–848 (2020).
11. Li, M.-Y., Li, L., Zhang, Y. & Wang, X.-S. *Infect. Dis. Poverty* **9**, 45 (2020).

AVISHEK DAS/SOPA IMAGES/LIGHTROCKET/GETTY

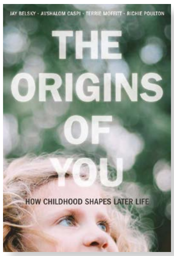
Children in India wait for a meal during the COVID-19 pandemic.

The lifelong studies that hold clues for kids' futures

Decades of data on how childhood affects adult health should help policymakers to plan. **By Barbara Maughan**

The COVID-19 pandemic has disrupted the lives of children around the world. How will this once-in-a-century event shape their development and later years? Biologists and social scientists have some ideas, thanks to a growing body of empirical evidence from long-term research on cohorts of people recruited at birth and studied regularly over decades, with some participants now in their seventies. This work has revealed, for instance, that low birth weight is associated with an increased risk of high blood pressure many decades later, and that level of education has implications for life expectancy. Such findings have shaped early-years interventions in many nations.

Four leaders in cohort studies share insights from their own work in *The Origins of You* (written before the pandemic). Psychologists Jay Belsky, Avshalom Caspi, Terrie Moffitt and Richie Poulton have between them set up and run three remarkable projects in New Zealand, the United States and the United Kingdom, tracking children from birth into their teens, twenties, thirties or forties. Every few years, participants are assessed on everything from their height, weight and impulsivity to their school results, pay, personality and mood. The authors hope to convey enthusiasm beyond academia for their adventures in science.



The Origins of You: How Childhood Shapes Later Life
Jay Belsky, Avshalom Caspi, Terrie E. Moffitt & Richie Poulton
Harvard Univ. Press (2020)

Twenty chapters cover examples of these adventures in conversational style, navigating increasingly complex ideas about the studies' concepts, methodologies and content. For example, Nobel-prizewinning economist James Heckman was keen to understand why participants in some Head Start programmes – launched in the 1960s to provide educational and health support for US children from low-income families – fared better in education and employment later in life, even though their early gains in test scores faded with time.

Heckman's hunch was that the long-term benefits might have come about because the programme had improved the children's self-control. He encouraged the authors to look into the matter. Sceptical, the team mined data from the Dunedin Multidisciplinary Health and Development Study, which has followed 1,000 New Zealanders since they were born in 1972–73. The researchers looked for indicators of level of self-control in childhood, and tested how well these predicted aspects of the study participants' later lives.

It turned out that Heckman was right. Even after controlling for factors such as family socio-economic status, worse self-control in childhood predicted a plethora of adverse outcomes: poorer physical health in the early thirties; lower social status and wealth; and increased risks of drug and alcohol use and being convicted of a crime. Importantly, these predictions showed a gradient across the range of early self-control, suggesting that strategies to enhance this quality – from behavioural 'nudges' to parent training programmes – would pay dividends, whatever the individual child's starting point.

Developmental profiles

Each chapter has its own childhood-to-adulthood story, and includes pointers for policy. I was especially struck by studies that focus on particular developmental periods, such as the impact that variations in the timing of puberty can have on early sexual behaviour in teenagers, and on the stability of sexual partnerships even up to a person's forties. Other chapters are designed to clarify the developmental profiles of conditions such as attention-deficit hyperactivity disorder (ADHD), and use data from across developmental periods to chart how they wax and wane with age. Similar analyses are conducted for antisocial behaviour.

Perhaps the most important theme that emerges is that although clear continuities exist between childhood and later well-being, these links are far from exact. Human development is probabilistic rather than deterministic, and continues well beyond the first decade of life. Many different processes are likely to underlie such long-term continuities. We see here, for example, instances of the ways in which childhood characteristics can 'select' individuals into later environments, so reinforcing early tendencies. Tracked to early adulthood, for example, people who were socially inhibited as toddlers had smaller social circles and less social support than their peers, whereas those who had been impulsive in early childhood often evoked negative responses from family, friends and partners, and in the workplace. Early adversities such as maltreatment, social isolation and bullying can become embedded in our biology, influencing inflammatory processes and stress responses in ways that might, later in life, increase the risk of conditions such as diabetes and poor mental health.

But we also see that change is possible throughout life, and that some individuals are resilient even in the face of quite severe early adversity. Teasing out the factors that contribute to strengths, whether they lie in the family, the neighbourhood, society or genetic inheritance, can be especially valuable in pointing to targets for intervention — such as investment in school meals or education.

Long game

Alongside the specific findings, what shines through is the power of the longitudinal method. The three projects that are explored here form part of a larger body of studies, mostly initiated since the Second World War, tracking individuals across their lives. Their findings are now revolutionizing our understanding of the determinants of health and social capital, and, in the case of the longest-running studies, of ageing and decline. Each represents an extraordinary investment — by researchers, participants and, of course, funders — in documenting lives in real time.

A child takes part in a 1940 development study.

HANSEL MIETH/THE LIFE PICTURE COLLECTION/GETTY

It's true that essentially 'observational' studies might not give the tight purchase on causality that could be achieved by an experiment. Instead, they offer something in many ways richer and more valuable: insights into the processes that shape human development. Given the tricks that memory can play, issues of this kind cannot be studied retrospectively. We need to observe lives as they unfold. And as this book shows, the value of such data increases exponentially with time, illuminating issues undreamt of when the studies began.

For those new to cohort literature, *The Origins of You* is an engaging introduction. For those familiar with this work, it is a chance to hear the authors thinking aloud, debating the best approaches and pondering what to study next. We can be certain that those conversations will now include how best to use these rich longitudinal resources to understand the effects of COVID-19.

Barbara Maughan is professor of developmental epidemiology at the Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience at King's College London.
e-mail: barbara.maughan@kcl.ac.uk

The poisonous history of chemotherapy

A Second World War disaster drove a crusade for cancer treatment argues Jennet Conant. **By Heidi Ledford**

On 2 December 1943, German forces attacked the Italian port town of Bari. The onslaught cost at least 1,000 lives and sunk 17 ships. One was carrying 2,000 bombs loaded with deadly mustard gas.

The gas — which was actually in liquid form — mixed with oil from the sinking tankers to create a deadly slick that clung to sailors' skin as they swam to safety. Many who made it to the local hospital were greeted with blankets to wrap around their poison-soaked clothing, sealing their fate as they awaited care. The agony set in hours or days later. Stunned nurses found themselves with wards full of swollen, blistered patients, temporarily blinded.

The Great Secret brings that harrowing night to life, and then follows the military physician who fought to uncover the truth about the chemical weapons. His efforts contributed to the development of chemotherapy, seeding the cancer-research juggernaut that dominates drug discovery to this day, argues writer Jennet Conant in her latest history of war-era science.

That hard-working and brilliant physician is the first of the book's two heroes. Stewart Alexander, an American expert on chemical weapons, is called in to explain the mysterious ailments plaguing the Bari survivors. The possibilities offer a harrowing tour through the chemical arms race of the early twentieth century. Could it have been chlorine or mustard, the causes of the chemical massacres of the First World War? Or was it lewisite, a blistering agent that quickly penetrated the skin? Or one of the new blends such as 'Winterlost', a combination of nitrogen mustard and lewisite that featured a low freezing point to ensure effectiveness at the frigid Russian front?

Chemical secret

The deadly cargo in Bari's harbour was a fiercely guarded secret. The Geneva Protocol had banned the use of chemical warfare in 1925, but the shipment was there in case of the need to retaliate if Hitler had resorted to chemical weapons. Alexander struggles to treat his ailing patients while battling military officials who are intent on keeping the incident quiet.

Alexander is struck by how the mustard–oil mixture obliterated his patients' white blood cells. He scrambles to make sense of data from different treatments given in different hospitals, with different standards of care and no control groups. (There are uncomfortable parallels with the flurry of uninterpretable observational studies and uncontrolled clinical trials during the first months of the COVID-19 pandemic.)

Alexander had seen similar effects of such agents in animal studies before the war. These had conjured up hopes that the chemicals could be used to rein in cancerous blood cells in leukaemia and lymphoma. Flood the body with toxic substances, the theory went, and the disease could be snuffed out or at least beaten back. Alexander's detailed report of


MONDADORI/GETTY

Rescuers work sift through the debris left by the explosion of a munitions ship in Bari harbour, Italy, in 1945.

his findings in Bari, initially classified but circulated among some military researchers, spurred efforts to find a chemical treatment for cancer.

On this point, Conant has to labour to connect the dots. The inspiration for chemotherapy did not come from Bari. Yale University researchers in New Haven, Connecticut, first treated cancer with nitrogen mustard in 1942; the patient died of lymphosarcoma a year before the Germans attacked the Italian harbour. But Conant argues that Alexander's report of his observations helped to convince researchers of the value and robustness of the approach.

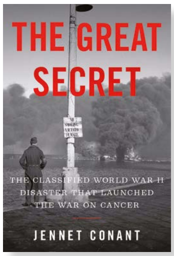
The book's second protagonist is physician Cornelius 'Dusty' Rhoads. He is much harder to like. Fiercely driven and passionate about curing cancer, Rhoads oversold preliminary research results and rushed into clinical trials. Before the war, Rhoads worked at Rockefeller University in New York City, and he travelled to Puerto Rico to study conditions such as anaemia and tropical sprue. There, he penned a hideously racist letter – unsent but discovered by his office staff – claiming to have transplanted cancer cells into healthy Puerto Ricans, whom he compared to animals. Rhoads later said the claim was a joke; subsequent investigations found no evidence that he carried out such "experiments".

Nevertheless, Rhoads continued to wield significant influence in military and academic science. He applied that influence with full force to the search for chemotherapies. Scepticism from other physicians was rampant. Cancer treatment, Conant reminds us, had changed little since Hippocrates (460–370 BC) named the disease and proclaimed "what drugs will not cure, the knife will". Surgery and radiation were nearly the only options, and cancer was so lethal and stigmatized that patients often were not told of their diagnosis.

Hope and heartbreak

After the war, Rhoads advocated fiercely for chemotherapy – inspired in part by Alexander's report, Conant argues. Rhoads's leadership and aggressive fundraising led, by the mid-1950s, to the first large-scale efforts to screen for new cancer drugs and to test promising candidates in people. Conant brings to life the exhilaration and hope that physicians felt when the first patients responded to chemotherapy – followed by the heart-wrenching dismay when, time and again, initial success was followed a few weeks or months later by the cancer's resurgence.

Opponents were horrified by the toxicity of chemotherapies and unimpressed by the ephemeral reprieves that most offered. US physician William Woglom captured the challenge: "It is almost, not quite, but almost as hard as finding some agent that will dissolve away the left ear, say, but leave the right ear unharmed; so slight is the difference between the cancer cell and its normal ancestor."



The Great Secret: The Classified World War II Disaster that Launched the War on Cancer
Jennet Conant
W. W. Norton (2020)

Despite that challenge, Rhoads planted the seeds for the cancer-research enterprise that continues today. There are now reams of DNA sequence data detailing the genetic differences between our 'left and right ears'. Drug-screening efforts are more sophisticated, and the chemical libraries that they trawl are orders of magnitude larger and more complex.

For a science-hungry reader, *The Great Secret* has a few too many excursions into the strategies, personalities and troop movements of the Second World War. And I yearned for more on the development of ethical boundaries between experimentation and treatment, which remain fuzzy in cancer research. But the book succeeds as a history of chemotherapy's origins.

Today, chemotherapy has advanced; some drugs are less toxic, given at lower doses, or more-targeted in their effects. But the benefits are still too often transient. "For a short period of time the patient was delighted," says one researcher of the first mustard chemotherapy trial. "But it was a short period of time."

Heidi Ledford is a senior reporter for *Nature* in London.

AFP VIA GETTY

Protests against racism in Detroit, Michigan, and many other US cities in 1967 prompted attempts to forecast future demonstrations.

Scientists use big data to sway elections and predict riots – welcome to the 1960s

Jill Lepore

A cold-war-era corporation targeted voters and presaged many of today's big-data controversies.

Ignorance of history is a badge of honour in Silicon Valley. "The only thing that matters is the future," self-driving-car engineer Anthony Levandowski told *The New Yorker* in 2018 (ref. 1).

Levandowski, formerly of Google, Uber and Google's autonomous-vehicle subsidiary Waymo (and recently sentenced to 18 months in prison for stealing trade secrets), is no outlier. The gospel of 'disruptive innovation' depends on the abnegation of history[2]. 'Move fast and break things' was Facebook's motto. Never look back. Another word for this is heedlessness. And here are a few more: negligence, foolishness and blindness.

Much of what technology leaders tout as original has been done before – and long ago. Yet few engineers and developers realize that they're stuck in a rut. That lack of awareness has costs, both economic and ethical.

Consider the strange trajectory of the Simulmatics Corporation, founded in New York City in 1959. (Simulmatics, a mash-up of 'simulation' and 'automatic', meant then what 'artificial intelligence (AI)' means now.) Its controversial work included simulating elections – just like that allegedly 'pioneered' by the now-defunct UK firm Cambridge Analytica on behalf of UK Brexit campaigners in 2015 and during Donald Trump's US presidential election campaign in 2016.

Journalists accused Trump's fixers of using a "weaponized AI propaganda machine" capable of "nearly impenetrable voter manipulation". New? Hardly. Simulmatics invented that in 1959. They called it the People Machine.

As an American historian with an interest in politics, law and technology, I came across the story of the Simulmatics Corporation five years ago when researching an article about the polling industry[3]. Polling was, and remains, in disarray. Now, it's being supplanted by data science: why bother telephoning someone to ask her opinion when you can find out by tracking her online?

Wondering where this began took me to the Massachusetts Institute of Technology (MIT) in Cambridge, to the unpublished papers of political scientist Ithiel de Sola Pool. He helped to establish the Simulmatics Corporation and led the cold-war-era campaign to bring behavioural science into the defence industry, campaigning and commerce. This story struck me as so essential to modern ethical dilemmas around data science, from misinformation and election interference to media manipulation and predictive policing, that I wrote a book about it: *If Then: How the Simulmatics Corporation Invented the Future* (2020).

Simulmatics, hired first by the US Democratic Party's National Committee in 1959 and then by the John F. Kennedy campaign in 1960, pioneered the use of computer simulation, pattern detection and prediction in American political campaigning. The company gathered opinion-poll data from the archives of pollsters George Gallup and Elmo Roper to create a model of the US electorate.

They split voters into 480 types – Democratic female blue-collar Midwesterner who voted for Democratic presidential candidate Adlai Stevenson in 1952 but for the Republican Dwight D. Eisenhower in 1956, say. And they assigned issues of concern, such as the importance of civil rights or a strong stand against the Soviet Union, into 60 clusters. It was, at the time, the largest such project ever conducted. It involved what Simulmatics called "massive data" decades before 'big data' became a buzzword.

Simulmatics was staffed by eminent scientists. Led by Pool, the group included researchers from MIT, Yale University in New Haven, Connecticut, Johns Hopkins University in Baltimore, Maryland, and Columbia University in New York City. It also included Alex Bernstein from IBM, who had written the first chess-playing computer program. Many of them, including Pool, had been trained by Yale political scientist Harold Lasswell, whose research on communication purported to explain how ideas get into people's heads: in short, who says what, in which channel, to whom, with what effect? During the Second World War, Lasswell studied the Nazis' use of propaganda and psychological warfare. When those terms became unpalatable after the war ended, the field got a new name – mass-communications research. Same wine, new bottle.

Like Silicon Valley itself, Simulmatics was an artefact of the cold war. It was an age obsessed with prediction, as historian Jenny Andersson showed in her brilliant 2018 book, *The Future of the World*. At MIT, Pool also proposed and headed Project ComCom (short for Communist Communications), funded by the US Department of Defense's Advanced Research Projects Agency (ARPA). Its aim, in modern terms, was to try to detect Russian hacking – "to know how leaks, rumors, and intentional disclosures spread" as Pool described it.

The press called Simulmatics scientists the "What-If Men", because their work – programming an IBM 704 – was based on endless what-if simulations. The IBM 704 was billed as the first mass-produced computer capable of doing complex mathematics. Today, this kind of work is much vaunted and lavishly funded. The 2018 *Encyclopedia of Database Systems* describes 'what-if analysis' as "a data-intensive simulation". It refers to it as "a relatively recent discipline". Not so.

Winning ways

John F. Kennedy won the 1960 US presidential election by the closest popular-vote margin since the 1880s – 49.7% to Richard Nixon's 49.5%. Before Kennedy's inauguration, a storm erupted when *Harper's* magazine featured a shocking story: a top-secret computer called the People Machine, invented by mysterious What-If Men, had in effect elected Kennedy. Lasswell called it "the A-bomb of the social sciences".

"Their very lack of interest in contemplating the possible consequences of their work stood as a terrible danger."

Kennedy had been trailing Nixon in the polls all summer. He had gained on Nixon in the autumn for three reasons: Kennedy championed civil rights and increased his share of African American votes; as a Catholic, he took a strong stance on freedom of religion; and he outperformed Nixon in four televised debates. Simulmatics had recommended each of these strategies.

Uproar broke out. *The New York Herald Tribune* called the People Machine Kennedy's "secret weapon". *The Chicago Sun-Times* wondered whether politicians of the future would have to "Clear it with the P.-M.". An Oregon newspaper expressed the view that Simulmatics had reduced voters to "little holes in punch cards", and that, by denying the possibility of dissent, the People Machine made "the tyrannies of Hitler, Stalin and their forebears look like the inept fumbling of a village bully".

Worse, Kennedy had campaigned against automation. In St Louis, Missouri, in September 1960 he'd delivered a speech warning about the "replacement of men by machines". A Kennedy campaign brochure asked: "If Automation takes over your job … who will you want in the White House?" Newspaper editors and commentators charged him with hypocrisy.

The ensuing debate raised questions that are still asked today – urgently. Can computers rig elections? What does election prediction mean for democracy? What does automation mean for humanity? What happens to privacy in an age of data? There were no answers then, as now. Lasswell merely admitted: "You can't simulate the consequences of simulation."

The most prescient critique came from another of Lasswell's former collaborators, Eugene Burdick. His dystopian novel *The 480*, published in 1964, described a barely fictionalized organization called Simulations Enterprises. In a sober preface, Burdick, a political scientist at the University of California, Berkeley, and bestselling novelist – known for co-authoring *The Ugly American* in 1958 – warned against the political influence of what is now called data science.

"The new underworld is made up of innocent and well-intentioned people," he wrote. Most of them are "highly educated, many with PhDs". They "work with slide rules and calculating machines and computers which can retain an almost infinite number of bits of information as well as sort, categorize, and reproduce this information at the press of a button".

Although none of the researchers he had met "had malignant political designs on the American public", Burdick warned, their very lack of interest in contemplating the possible consequences of their work stood as a terrible danger. Indeed, they might "radically reconstruct the American political system, build a new politics, and even modify revered and venerable American institutions – facts of which they are blissfully innocent".

Burdick knew these researchers, and he had worked with Pool as well as Lasswell. He spied in their ambition, in their enthralment with the capacities of computers, the wide-eyed heedlessness that remains Silicon Valley's Achilles heel.

Big business

Buoyed by the buzz of Kennedy's election, Simulmatics began an advertising blitz. Its 1961 initial stock offering set out how the company would turn prediction into profit – by gathering massive data, constructing mathematical models of behavioural processes, and using them to simulate "probable group behaviour".

The firm pitched its services to media companies, government departments and advertising agencies, with mixed success. It persuaded executives from the Motion Picture Association of America, MGM film studios and Columbia Records to set up forms of analysis that would ultimately, when it was possible to collect enough data to make this work, lead to Netflix and Spotify. It proposed a "mass culture model" to collect consumer data across all media – publishing houses, record labels, magazine publishers, television networks, and film studios – to direct advertising and sales. It sounds a lot like Amazon.

Simulmatics introduced what-if simulation to the advertising industry, targeting consumers with custom-fit messages. In 1962, it became the first data firm to provide real-time computing to a US newspaper, *The New York Times*, for analysing election results. For the government, it proposed models to aid public-health campaigns, water-distribution systems, and, above all, the winning of hearts and minds in Vietnam.

In 1963, on behalf of the Kennedy administration, Simulmatics simulated the entire economy of Venezuela, with an eye to halting the advance of socialism and communism. A larger project to undertake such work throughout Latin America, mostly designed by Pool and known as Project Camelot, became so controversial that the next president, Lyndon B. Johnson, dismantled it.

After 1965, Simulmatics conducted psychological research in Vietnam as part of a bigger project to use computers to predict revolutions. Much of this work built on earlier research by Lasswell and Pool, identifying and counting keywords, such as 'nationalism', in foreign-language newspapers that might indicate the likelihood of coups. Such topic-spotting is the precursor to Google Trends.

Growing unrest

Simulmatics brought those counter-insurgency methods home in 1967 and 1968, as protests against racial injustice broke out on the streets of US cities such as Los Angeles, California, and Detroit, Michigan. The company attempted to build a race-riot prediction machine for the Johnson administration. It failed. But its cockeyed ambition – the drive to forecast political unrest – was widely shared, and has endured, not least in the ethically indefensible work of predictive policing.

Civil-rights activists, then as now, had little use for such schemes. "I will not predict riots," James Farmer, head of the Congress of Racial Equality, said on CBS TV's *Face the Nation* in April 1965. "No one has enough knowledge to know that." The real issue, he pointed out, was that no one was addressing the problems that led to unrest. "I am not going to predict rioting here," Martin Luther King Jr told the press in Cleveland, Ohio, in June 1967.

But the fantasy of computer-aided riot prediction endured, as widely and passionately held as the twenty-first century's dream that all urban problems can be solved by 'smart cities', and that civil unrest, racial inequality and police brutality can be addressed by more cameras, more data, bigger computers and yet more what-if algorithms.

Predictive demise

Simulmatics began to unravel in 1969. Student protesters at MIT accused the company of war crimes for its work in Vietnam. They even held a mock trial of Pool, calling him a war criminal. "Simulmatics looks like nothing more than a dummy corporation through which Pool runs his outside Defense work," the *New Republic* reported. "Simulation companies are not so popular as they once were; their proprietors are often regarded as cultists, and the generals who were persuaded to hire them by liberals in the Kennedy and early Johnson administrations are sour on the whole business."

There were problems with early predictive analytics, too. Data were scarce, computers were slow. Simulmatics filed for bankruptcy in 1970, and vanished.

Pool went on to become a prophet of technological change. "By 2018 it will be cheaper to store information in a computer bank than on paper," he wrote in 1968, in a contribution to a book called *Toward the Year 2018* (ref. 4). Tax returns, social security and criminal records would all be stored on computers, which could communicate with one another over a vast international network.

People living in 2018 would be able to find out anything about anyone, he wrote, without ever leaving their desks. "The researcher sitting at his console will be able to compile a cross-tabulation of consumer purchases (from store records) by people of low IQ (from school records) who have an unemployed member of the family (from social security records)."

Would he have the legal right to do so? Pool had no answer: "This is not the place to speculate how society will achieve a balance between its desire for knowledge and its desire for privacy."

Collective amnesia

Before his early death in 1984, Pool was also a key force behind the founding of the most direct descendant of Simulmatics, the MIT Media Lab. Pool's work underlies the rules – or lack of them – that prevail on the Internet. Pool also founded the study of "social networks" (a term he coined); without it, there would be no Facebook. Pool's experiences with student unrest at MIT – and especially with the protests against Simulmatics – informed his views on technological change and ethics. Look forward. Never look back.

In 1966, Pool described the social sciences as "the new humanities of the Twentieth Century"[5]. Although leaders in times past had consulted philosophy, literature and history, those of the cold-war era, he argued, were obligated to consult the social sciences. Given a choice between "policy based on moralisms and policy based on social science", he was glad to report that the United States, in conducting the war in Vietnam, had rejected the former in favour of rationality.

To me, this sounds a lot like Levandowski. "I don't even know why we study history," Levandowski said in 2018 (ref. 1). "It's entertaining, I guess – the dinosaurs and the Neanderthals and the Industrial Revolution and stuff like that. But what already happened doesn't really matter." Except, it does matter. Attempting to thwart revolt and defeat social unrest by way of predictive algorithms has been tried before; it failed, and was ethically indefensible.

This summer, under pressure from the Black Lives Matter movement, US police departments are abandoning predictive policing, an industry led by the data-analytics firm PredPol in Santa Cruz, California. IBM and Google have, at least publicly, pulled back from another form of algorithm-driven surveillance, facial recognition. Maybe these detours might have been avoided if the people developing them had stopped to consider their origins in the Vietnam War.

It's worth remembering, too, that protesters at the time understood that connection. In 1969, MIT activists objecting to companies such as Simulmatics asked what, really, was the point of making human behaviour a predictive science, in a world of agonizing inequalities of power. What was it all for? How was it likely to be used?

As one student protester asked in an anti-war pamphlet: "To do what? To do things like estimate the number of riot police necessary to stop a ghetto rebellion in city X that might be triggered by event Y because of communications pattern K given Q number of political agitators of type Z?"

It's a question worth asking today, all over again.

The author

Jill Lepore is professor of American history at Harvard University in Cambridge, Massachusetts, a staff writer at *The New Yorker*, host of podcast The Last Archive, and author of *If Then: How the Simulmatics Corporation Invented the Future* (2020).

1. Duhigg, C. 'Did Uber Steal Google's Intellectual Property?' *The New Yorker* (22 October 2018).
2. Lepore, J. 'The Disruption Machine' *The New Yorker* (23 June 2014).
3. Lepore, J. 'Politics and the New Machine' *The New Yorker* (16 November 2015).
4. Mesthene, E. G. (ed.) *Toward the Year 2018* (Foreign Policy Association, 1968).
5. de Sola Pool, I. *Backgr.* **10**, 111–122 (1966).

Expert insight into current research

News & views

Metabolism

Light-activated neurons can alter body heat

Gary J. Schwartz

A light-sensitive receptor protein expressed in neurons deep in the mouse brain has been shown to be stimulated by violet light, and to activate a pathway that reduces heat production in brown fat. **See p.420**

Light has profound effects on human behaviour and physiology, from synchronizing sleep–wake cycles to inducing daily fluctuations in body temperature and energy metabolism. Our ability to see is mediated by a family of opsin proteins in the retina. When exposed to light, opsins modulate the flow of ions across neuronal membranes, ultimately activating the optic nerves[1]. In mammals, an opsin called opsin 5 (OPN5) is expressed in an unusual place – in neurons deep in the brain's preoptic area[2] (POA), which has a role in metabolism. On page 420, Zhang *et al.*[3] report a pathway by which OPN5 in the POA regulates heat production in mice. The authors' findings open up the possibility of modulating metabolism by manipulating environmental light.

Zhang and colleagues first asked which neurons activate OPN5-containing POA cells in mice. They injected the POA with a tracer virus that selectively labels OPN5-containing neurons. The tracer is taken up by the nerves that send impulses to these cells, by the nerves that feed into them, and so on up the neuronal circuit. The authors found that OPN5-containing neurons receive input from multiple pools of neurons in the forebrain and brain stem.

These upstream neurons are all part of a circuit that senses changes in skin temperature and controls regulatory responses in a type of fat called brown adipose tissue (BAT). The main role of BAT is to generate heat, raising body temperature as it burns fuel. Heat production is stimulated by the neurotransmitter noradrenaline, which is released from neurons of the sympathetic nervous system in response to cold temperatures. Noradrenaline binds to β_3-adrenergic-receptor proteins on the brown-fat cells, rapidly triggering fuel burning and robust heat production.

Zhang *et al.* next injected a tracer virus into the BAT. The tracer labelled the entire circuit of neurons upstream of the BAT, and confirmed that the OPN5-expressing neurons are part of the circuit that projects into BAT. The group found that these neurons express three neurochemicals: glutamate, pituitary adenylate cyclase-activating peptide and brain-derived neurotrophic factor. This combination has previously been shown to be characteristic of heat-sensitive neurons[4].

The authors modulated the activity of the OPN5 neurons by engineering them to express synthetic excitatory or inhibitory ion-channel proteins, which, respectively, activate or inhibit neurons in response to an injected chemical. Stimulation of the excitatory channels rapidly and robustly reduced heat production by BAT, and so reduced core body temperature. These data indicate that the OPN5 neurons inhibit BAT activity (Fig. 1). By contrast, stimulation of the inhibitory channels increased core temperature.

In line with these results, mice engineered to lack the *Opn5* gene showed higher BAT activity and body temperature than did controls. They also exhibited a raft of other metabolic changes: increased energy expenditure, smaller fat cells, lower fat-pad weights, lower levels of circulating cholesterol and better resistance to environmental cold.

OPN5 responds to violet light, and Zhang and colleagues found that the mutant mice were insensitive to violet light. By contrast, violet light induced a decrease in BAT activity and core temperature in control animals. The authors also raised control animals in the absence of violet light throughout embryonic and postnatal development. Under such lighting, these mice were resistant to environmental cold, similar to animals lacking OPN5.

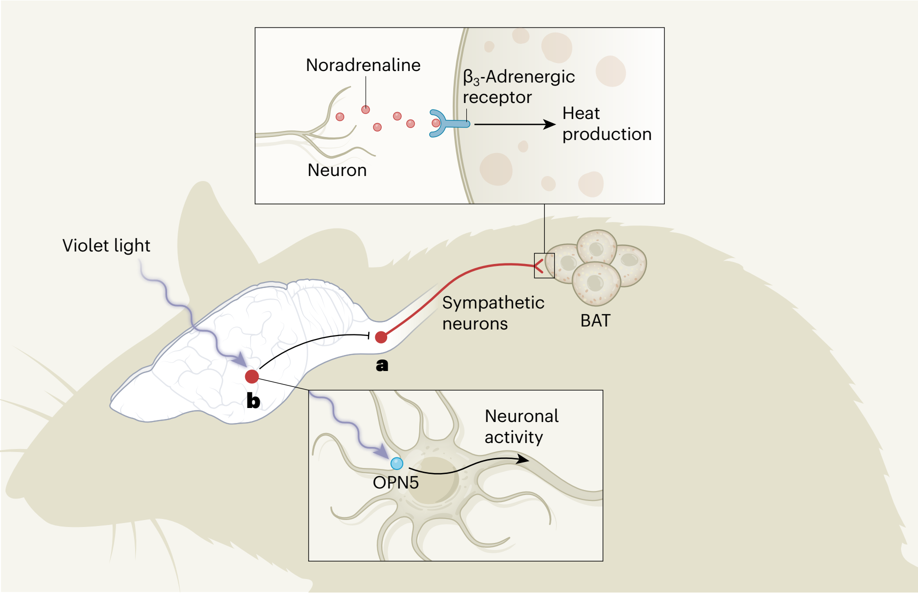


Figure 1 | Shining a light on heat production in mice. a, Neurons of the sympathetic nervous system project from the brain to the cells of brown adipose tissue (BAT). These neurons release the neurotransmitter molecule noradrenaline, which binds to β_3-adrenergic-receptor proteins on the BAT cells, triggering the cells to break down glucose and so produce heat. **b**, Zhang *et al.*[3] report that violet light activates a light-sensitive protein called opsin 5 (OPN5) on neurons in the preoptic area of mouse brains. When activated, these neurons inhibit the pathway outlined above, and so prevent heat production.

It is important to note that the *Opn5*-mutant animals did not express the gene at any point in their lives – including during crucial developmental periods when the neural circuitry and identities of neurons are established. It is not yet known whether this led to unexpected developmental changes that might underlie the animals' insensitivity to violet light. Going forward, the same analysis should be performed in animals in which *Opn5* is deleted only during adulthood, after normal neurological development has finished.

To prove that violet light could penetrate the skull and reach the POA neurons, Zhang *et al.* implanted a miniature, wavelength-sensitive radiometer probe into the brain. They found that violet light could indeed penetrate deep enough to activate OPN5-expressing POA neurons. Finally, they compared the response to cold of animals exposed to a full spectrum of light and of animals exposed to light that lacked violet wavelengths. The 'full-spectrum' animals showed greater reductions in BAT and body temperature in response to cold than did the 'minus-violet' animals. This experiment indicates a physiologically relevant role for OPN5-expressing POA neurons – repressing heat production in BAT in response to violet light.

Whether violet light directly stimulates OPN5 neurons remains to be proved. Zhang *et al.* used neuroimaging techniques to show that light activates the neurons in tissue slices, but proof will involve applying these techniques *in vivo*.

OPN5 has been identified in the hypothalamus (the brain region in which the POA is located) in monkeys[2]. However, we do not yet know whether ambient light will reach this deep brain region. Such a demonstration would be a key step in determining the applicability of these results to humans.

As with many exciting and unanticipated findings, Zhang and colleagues' study opens the door to larger questions of biological relevance. Humans today have unprecedented control over ambient light, temperature and nutrient supply, and are consequently much less susceptible to natural environmental metabolic challenges than were our ancestors. Eating only during daylight hours has been shown to markedly improve insulin sensitivity in people with prediabetes[5] – a change that might lower the risk of developing full-blown diabetes. It is tempting to speculate that limiting violet light might activate BAT, and thereby augment the metabolic benefits of daytime-restricted eating. Similarly, drugs called β-agonists activate BAT, lower blood glucose levels and increase resting metabolic rate and insulin sensitivity in people[6–9], and Zhang and co-workers demonstrated that animals reared without violet light show increased responses to these drugs. Limiting violet light might therefore extend the beneficial metabolic effects of β-agonists.

Remarkably, mouse and human BAT expresses a red-light-sensitive protein, OPN3 (ref. 9). Red-light stimulation of OPN3 increases glucose uptake and heat production in BAT, both *in vitro* and in mice. Thus, different spectra of environmental light might act both in the brain and in brown-fat cells to alter BAT heat production in ways that can help the body to control glucose levels.

Finally, a population of neurons has recently been found in the mouse POA that controls torpor – a state characterized by low body temperature and a markedly reduced metabolic rate, typically induced by harsh environmental challenges such as cold and lack of food[10]. It remains an open question whether this neuronal circuit is also sensitive to violet light. But Zhang and colleagues' findings raise the possibility that environmental light might orchestrate a host of coordinated brain responses that together determine the highs and lows of metabolism.

Gary J. Schwartz is at the Fleischer Institute for Diabetes and Metabolism, and in the Departments of Medicine, of Neuroscience, and of Psychiatry and Behavioral Sciences, Albert Einstein College of Medicine, New York, New York 10461, USA.
e-mail: gary.schwartz@einsteinmed.org

1. Milosavljevic, N. *et al. Proc. Natl Acad. Sci. USA* **115**, E11817–E11826 (2018).
2. Yamashita, T. *et al. J. Biol. Chem.* **289**, 3991–4000 (2014).
3. Zhang, K. X. *et al. Nature* **585**, 420–425 (2020).
4. Tan, C. L. *et al. Cell* **167**, 47–59 (2016).
5. Sutton, E. F. *et al. Cell Metab.* **27**, 1212–1221 (2018).
6. Finlin, B. S. *et al. J. Clin. Invest.* **130**, 2319–2331 (2020).
7. O'Mara, A. E. *et al. J. Clin. Invest.* **130**, 2209–2219 (2020).
8. Cypess A. M. *et al. Cell Metab.* **21**, 33–38 (2015).
9. Sato, M. *et al. PLoS Biol.* **18**, e3000630 (2020).
10. Hrvatin S. *et al. Nature* **583**, 115–121 (2020).

Developmental biology

Keratin as an aide-memoire

Mateusz Trylinski & Buzz Baum

Filaments of keratin – stable protein polymers best known for their function in hair and nails – provide a memory of cell polarity at a crucial stage in early mouse development. **See p.404**

The processes by which a single cell – the fertilized egg – gives rise to all the different cell types that make up an adult organism remain some of life's great mysteries. We know that it takes time for cells in an embryo to settle on a fate, because a single embryo that splits during early development can give rise to twins, triplets and more. But how are cell-fate decisions made, and how do cells coordinate their choices with their peers? Researchers have suggested numerous mechanisms that influence the paths taken by cells in early mammalian embryos. On page 404, Lim *et al.*[1] describe a surprising role for a protein polymer, keratin, in the first of these decision-making processes.

> **"Keratins provide a physical memory of polarity that is relatively independent of cell-division events."**

Two of the main challenges of early development are to increase the number of cells through repeated rounds of cell division, and to ensure that these cells assume distinct forms and functions at the right time and place to generate functional tissues and organs. The two processes can be coupled through 'asymmetric' cell divisions. These are divisions that give rise to two sibling cells with distinct identities, either as a result of the asymmetric segregation of material, or in response to local differences in the extracellular environment that the cells encounter after division.

It is during the 8- to 16-cell transition that cells in early mammalian embryos first become asymmetrically organized – with subsets of proteins becoming concentrated at opposite cell poles, a feature called apical–basal polarity. Cell identity remains plastic at this stage, but daughter cells that end up at the periphery (termed the trophectoderm) of the 16-cell embryo give rise to the placenta, whereas daughter cells that end up inside the embryo contribute to the fetus.

The observation of apical–basal polarity at the 8- to 16-cell transition led to the proposal that the future identity of these cells is determined by the asymmetric inheritance of the outward-facing apical domain[2], which is rich in

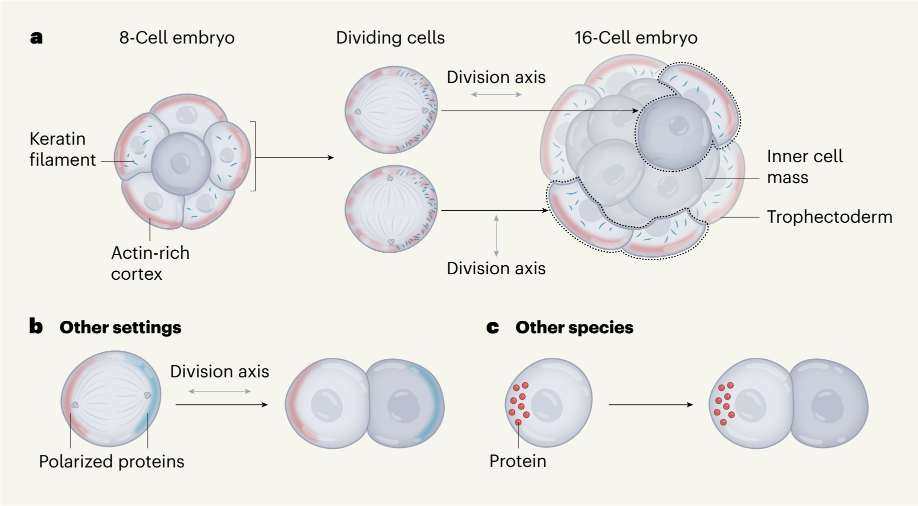


Figure 1 | Keratin in early embryos. a, In 8-cell mouse embryos (not all cells shown), keratin filaments are expressed stochastically in a subset of cells, before associating with a 'cortex' on one side of the cell (called the apical side). The cortex is rich in the protein actin. Lim and colleagues[1] show that, as these cells divide, the cortex disassembles but keratin filaments remain apically localized. The identity of daughter cells is determined by the position of keratin relative to the axis along which division occurs. Daughter cells that lack keratin end up inside the embryo and go on to form the inner cell mass. Daughter cells that inherit the mother's apical region also inherit keratin, which helps to re-establish the cortex at the apical pole. These cells contribute to the trophectoderm, from which the placenta arises. **b**, In most other settings studied, asymmetric division involves coordination between multiple polarized cues (such as polarized proteins) and a specific division axis. **c**, In other species, such as flies, asymmetric inheritance of protein aggregates provides a memory of the mother cell's state.

actin (a component of the cell's 'skeleton') and polarity proteins[3]. But, using fast live imaging, the group that performed the current study showed previously[3] that the apical domain is transiently lost during mitotic cell division, before re-forming in the daughter cells on the embryo's periphery. This puzzling observation suggested the existence of other factors that act as a memory of polarity during divisions. Following up hints from the old literature on mouse embryos[4], Lim and colleagues have now homed in on keratin, a type of intermediate filament protein.

Imaging keratin, the team observed a few short keratin polymers in a subset of cells in the 8-cell embryo. As these filaments grew during the part of the cell cycle between divisions, called interphase, they became preferentially associated with the apical, actin-rich cortex — a layer of proteins just inside the cell membrane. When the apical domain became disassembled during mitosis, these keratin filaments remained in place (Fig. 1a).

Although this might seem unexpected, other intermediate filaments have been shown to remain associated with the cortex during mitosis[5,6]. Lim *et al.* found that apical retention of these polymers depends on their slow diffusion, which is limited by their large molecular weight and the cytoplasmic actin meshwork in which they are embedded. As a result, keratin filaments are inherited by daughter cells that retain an outward face. So, once positioned at one end of the cell, these relatively inert stable polymers act as a physical memory of polarity.

The authors went on to show that, as cells of the new 16-cell embryo exit mitosis, inherited keratin filaments accelerate the repolarization of the apical cell cortex, which biases the cell towards becoming trophectoderm (through signalling pathways that involve Yap and Hippo proteins[7]). In turn, this bias is associated with high levels of keratin expression. So, over a period of hours, positive feedback in the system reinforces the accumulation of keratin in peripheral cells, and inhibits its expression in cells at the embryo's centre. By the 32-cell stage, when cell fate is more firmly established, the embryo itself is clearly polarized, with an outer, keratin-rich supporting cell layer, and inner cells that lack keratin.

Given its well-established role in stiffening epithelial cells[8], in an embryonic context keratin might both prevent outer cells from becoming internalized by apical constriction and help to give the trophectoderm its near-perfect spherical shape. Conversely, keeping keratin levels low in cells in the centre might help them retain the flexibility in shape that they require to generate a multilayered embryo.

By using keratin filaments to stably mark the peripheral cortex, mammalian embryos (in which patterns of cell division differ widely between individuals) can ensure that cells fated to become trophectoderm are always formed in the outer layer of the cell cluster, irrespective of the orientation of divisions. Keratins play a part as asymmetrically inherited fate determinants only in these relatively rare 'inside-out' divisions. The early mammalian embryo therefore differs from most other systems in which asymmetric division has been studied (Fig. 1b). In those cases, in order to impose a reproducible division asymmetry, the mitotic apparatus itself is oriented so that daughter cells inherit different complements of cortically localized cell-fate determinants[9].

In the coming years, it will be important to reconcile Lim and colleagues' data with suggestions of roles for the unequal segregation of messenger RNA encoding the Cdx2 protein[10] (one function of which is in forming the trophectoderm), or for differential contractility of the actomyosin protein complex[11], in the symmetry-breaking events that occur at this stage in mouse embryos. The fate of dividing cells that do not express keratin at the 8-cell stage also remains to be studied.

Taking a broader perspective, this work shows how the cellular function of a protein such as keratin can emerge from its physical characteristics. In early mouse embryos, keratins provide a physical memory of polarity that is relatively independent of cell-division events. In other organisms, from bacteria to multicellular animals, other proteins that polymerize or form aggregates have also been found to provide a physical memory of cell state during asymmetric divisions[12] (Fig. 1c). So Lim and co-workers' study provides another intriguing example of nature exploiting the material properties of a protein.

Mateusz Trylinski and **Buzz Baum** are in the MRC Laboratory for Molecular Cell Biology, University College London, London WC1E 6BT, UK.
e-mail: b.baum@ucl.ac.uk

1. Lim, H. Y. G. *et al. Nature* **585**, 404–409 (2020).
2. Johnson, M. H. & Ziomek, C. A. *Cell* **24**, 71–80 (1981).
3. Zenker, J. *et al. Cell* **173**, 776–791 (2018).
4. Jackson, B. W. *et al. Differentiation* **17**, 161–179 (1980).
5. Serres, M. P. *et al. Dev. Cell* **52**, 210–222 (2020).
6. Duarte, S. *et al. Nature Commun.* **10**, 4200 (2019).
7. Nishioka, N. *et al. Dev. Cell* **16**, 398–410 (2009).
8. Yamada, S., Wirtz, D. & Coulombe, P. A. *J. Struct. Biol.* **143**, 45–55 (2003).
9. Knoblich, J. A. *Nature Rev. Mol. Cell Biol.* **11**, 849–860 (2010).
10. Skamagki, M., Wicher, K. B., Jedrusik, A., Ganguly, S. & Zernicka-Goetz, M. *Cell Rep.* **3**, 442–457 (2013).
11. Maître, J.-L. *et al. Nature* **536**, 344–348 (2016).
12. Otzen, D. & Riek, R. *Cold Spring Harb. Perspect. Biol.* **11**, a033860 (2019).

Astronomy

A planet transiting a stellar grave

Steven Parsons

Evidence has been found of a planet circling the smouldering remains of a dead star in a tight orbit. The discovery raises the question of how the planet survived the star's death throes – and whether other planets also orbit the remains. **See p.363**

In the past few decades, the number of planets discovered beyond our Solar System has increased rapidly, and current estimates are that around one-third of all Sun-like stars host planetary systems[1]. Given that the Milky Way contains around ten billion Sun-like stars, there are likely to be billions of planets in our Galaxy. All of these planet-hosting stars will eventually die, leaving behind burnt-out remnants known as white dwarfs. What becomes of the stars' planetary systems when this happens is unclear, but in some cases it is thought that planets will survive and remain in orbit around the white dwarf[2]. On page 363, Vanderburg *et al.*[3] report the discovery of a planet that passes in front of (transits) the white dwarf WD 1856+534 every 1.4 days. Their work not only proves that planets can indeed survive the death of their star, but might offer us a glimpse of the far future of our own Solar System.

Sun-like stars fuse hydrogen into helium in their cores, producing copious amounts of energy that they use to support themselves against gravitational collapse. Stars are born with huge reserves of hydrogen, but eventually this supply is exhausted. The Sun has burnt through roughly half of its hydrogen supply. When this runs out, in five billion years, the Sun – and, by extension, the rest of the Solar System – will undergo a fundamental change.

When only a small amount of hydrogen remains, fusion will continue in a shell around the Sun's core. This will cause the outer envelope of the Sun to swell to an enormous size. At its maximum extent, the surface of the Sun might reach all the way to Earth's orbit, engulfing Mercury, Venus and, potentially, Earth itself. The Sun will then start to rapidly eject its outer envelope into interstellar space. The decreasing mass of the Sun will cause the other planets to move outwards, away from the Sun, to conserve angular momentum. When the last of the envelope is ejected, the Sun's core will be revealed: a smouldering, Earth-sized white dwarf that will slowly cool for the rest of time.

In this scenario, it is clear that the closest planets to the Sun are likely to be engulfed and destroyed. However, Mars, the asteroid belt and all the gas-giant planets will probably survive and stay in altered orbits around the Sun's remains. More broadly, we might expect many white dwarfs to host remnant planetary systems. Indeed, there has been growing evidence of this in the form of asteroids that have wandered too close to white dwarfs and then been torn apart by intense gravitational forces[4]. Debris from these asteroids rains down onto the surfaces of many white dwarfs, whereupon we can detect it[5]. However, until now, no planet in orbit around a white dwarf had been detected directly.

Enter Vanderburg *et al.*, who used data collected by NASA's Transiting Exoplanet Survey Satellite (TESS) mission to detect the periodic dimming of the white dwarf WD 1856+534. This dimming is caused by a planet passing between the white dwarf and Earth. Because white dwarfs are so small, the planetary transit is very 'deep': 56% of the white dwarf's light is blocked, compared with the typical 1–2% that is blocked by gas-giant planets around normal stars. In the case of WD 1856+534, the transiting planet is similar in size to Jupiter, and therefore has a diameter about ten times that of the white dwarf (Fig. 1).

In principle, such a deep transit should be easy to detect, so it might seem odd that such systems have escaped discovery for so long. However, the small size of white dwarfs also means that the transits are brief, lasting just 8 minutes in this case (compared with several hours for normal stars). Therefore, finding these planets requires white dwarfs to be both rapidly and constantly monitored – something that has become possible only in the past decade, thanks to missions such as TESS and the European Space Agency's (ESA's) Kepler (see ref. 6, for example).

The shape of the transit of WD 1856+534 gives us a good idea of the radius of the orbiting planet, but Vanderburg *et al.* were unable to place strong constraints on the planet's mass. Using infrared data, they calculate an upper limit of 14 times the mass of Jupiter. This confirms that the orbiting object is indeed a planet (rather than a failed star), but the unknown mass makes it impossible to tell whether the planet has been fundamentally

> **"Until now, no planet in orbit around a white dwarf had been detected directly."**

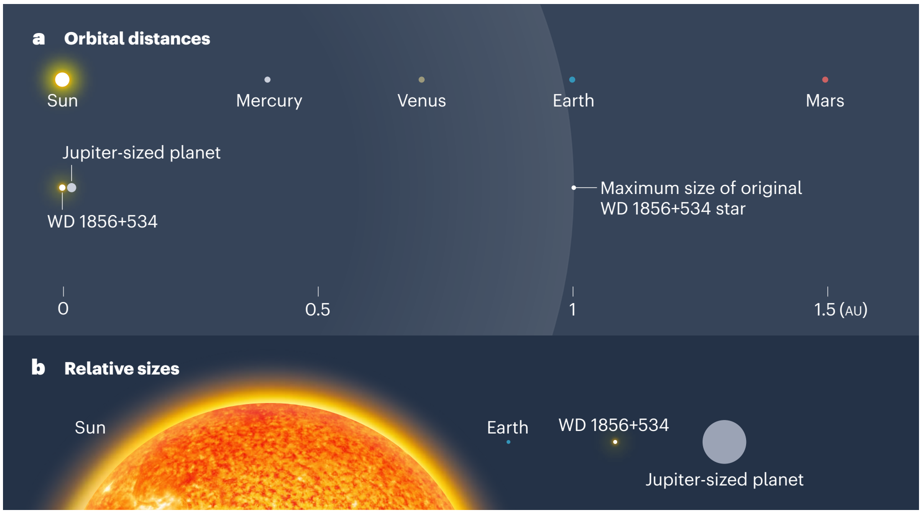


Figure 1 | Comparison between the inner Solar System and a white-dwarf system. Vanderburg *et al.*[3] report that a Jupiter-sized planet orbits the white dwarf WD 1856+534. **a**, The orbit is extremely small – the planet is roughly 20 times closer to the white dwarf than is Mercury to the Sun. The white dwarf was previously a giant star, the outer envelope of which once extended well beyond the planet's orbit. This raises the question of how the planet arrived in its current orbit. All distances are in astronomical units (AU), and the size of the giant star is shown to scale; the sizes of the other stars and planets are not shown to scale. **b**, The relative sizes of the Sun and Earth, and of WD 1856+534 and its orbiting planet, are shown here for comparison.

altered by the death of its host star. A mass and radius measurement for this planet would enable us to compare it with similar planets orbiting Sun-like stars, possibly revealing any changes that the planet has undergone in the past. Unfortunately, it seems highly unlikely that the mass will be determined precisely any time soon. This is because WD 1856+534 is too cold to produce any absorption features in its spectrum that could be analysed to determine the white dwarf's radial velocity, a measurement that is typically used to calculate the masses of orbiting planets.

One of the biggest questions to emerge from Vanderburg and colleagues' study is how the planet ended up so close to the white dwarf. The planet is located just 4 solar radii from the white dwarf (or roughly 20 times closer to the white dwarf than Mercury is to the Sun). Assuming that the inner planetary system was swallowed by the expanding star, it seems extremely unlikely that the planet has always been this close to its star.

Vanderburg *et al.* suggest two possible explanations. The first is that the planet avoided destruction by tearing off the outer layers of the expanding star when it was engulfed. The second is that several distant planets survived the death of the star, but their altered orbits caused them to interact with each other — whereupon the observed planet was thrown towards the white dwarf by another planet. This latter explanation seems the most likely, and offers the tantalizing prospect of detecting additional planets in this system in the future. Given that WD 1856+534 is only 25 parsecs (82 light years) from Earth, the gravitational effects of any further planets on the white dwarf could be detectable by missions such as ESA's Gaia space observatory. This system therefore opens up an entirely new field of exoplanetary research.

Steven Parsons is in the Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK.
e-mail: s.g.parsons@sheffield.ac.uk

1. Zhu, W., Petrovich, C., Wu, Y., Dong, S. & Xie, J. *Astrophys. J.* **860**, 101 (2018).
2. Veras, D., Mustill, A. J., Bonsor, A. & Wyatt, M. C. *Mon. Not. R. Astron. Soc.* **431**, 1686–1708 (2013).
3. Vanderburg, A. *et al. Nature* **585**, 363–367 (2020).
4. Jura, M. *Astrophys. J. Lett.* **584**, 91–94 (2003).
5. Gänsicke, B. T. *et al. Mon. Not. R. Astron. Soc.* **424**, 333–347 (2012).
6. Hermes, J. J. *et al. Mon. Not. R. Astron. Soc.* **468**, 1946–1952 (2017).

Tumour biology

How cancer invasion takes shape

Karolina Punovuori & Sara A. Wickström

Skin cancers resulting from distinct mutations have characteristic tissue forms and different disease outcomes. Analysing the architecture of benign and aggressive tumours reveals how mechanical forces drive these patterns. **See p.433**

The interplay between form and function is a cornerstone of biology, and the dismantling of normal tissue organization is a hallmark of many diseases. A long-standing question is whether changes in tissue architecture are merely a by-product of destructive diseases such as cancer, or whether they actively influence disease progression. Distinct types of skin cancer are driven by specific genetic abnormalities and give rise to distinctive tumour shapes. However, how these structures arise, and whether their specific forms affect the different outcomes of benign and malignant cancers, has been unclear. On page 433, Fiore *et al.*[1] report an analysis of skin cancer in mice that uncovers some of the key principles involved.

The skin's outer region, called the epidermis, is made of layers of epithelial cells. Down in the basal layer at the bottom of the epidermis, stem cells divide to self-renew their population and to generate cells of the suprabasal layers above, each layer of which represents a further-differentiated state. The final stage of differentiation generates a layer of dead cells on the skin's surface, which are continually shed. The constant need to replace these dying cells creates high demand for the basal stem cells to divide and produce differentiated cells. Owing to their potency and long lifetime, these stem cells, which frequently acquire cancer-causing mutations, are the cells of origin for two common types of skin cancer. One is basal cell carcinoma (BCC), a benign tumour that does not usually spread into other tissues, and the second is squamous cell carcinoma (SCC), which is more aggressive and invasive[2,3].

Fiore and colleagues engineered mouse embryonic skin cells to express cancer-causing mutations. A mutation in the gene *SmoM2* that activates the Sonic Hedgehog signalling pathway produced 'budding' skin conformations, characteristic of BCC (Fig. 1). By contrast, a mutation in the gene *HRas* that causes hyperactivity in the RAS–MAPK pathway generated skin 'folds' similar to those found in SCC. Both types of mutation caused cancer cells to proliferate faster than did their surrounding normal cells, but the mechanical properties of the tumour environment differed profoundly between the two tumour types.

Using an impressively broad selection of methods and combining theoretical and experimental approaches, Fiore *et al.* demonstrated that the two cancer-promoting mutations had different effects on the production, turnover and stiffness of the basement membrane. This is a thin layer of specialized extracellular matrix material that separates the epidermal cells from the rest of the skin, such as the adjacent compartment below called the dermis. The authors report that the BCC-like tumours actively produced and remodelled the basement membrane, and the resulting extracellular matrix had low stiffness and was malleable in its response to forces generated by the cancer cells. By contrast, the SCC-like cells produced less basement membrane, and the absence of remodelling made the underlying extracellular matrix comparatively stiffer.

As the BCC-like tumour expanded, the compressive forces exerted by the rapidly dividing and thus crowded pool of cancer cells caused buckling of the epidermis and basement membrane, resulting in the growth of tumour buds. However, in SCC-like tumours, the same type of force generated by proliferation and cellular crowding exerted towards the stiffer basement membrane did not result in such tissue deformation, and instead the tumour formed wave-like folds. Importantly, Fiore and colleagues report that experimentally altering the basement membrane to mimic high remodelling forced a switch from the formation of tumour folds to buds.

The authors observed specific differences between the two tumour types in the distribution of the actin and myosin protein machinery that generates cellular contractility and tension: the BCC-like cells exhibited high tension at the cellular boundary between the cancer and the neighbouring healthy tissue, however,

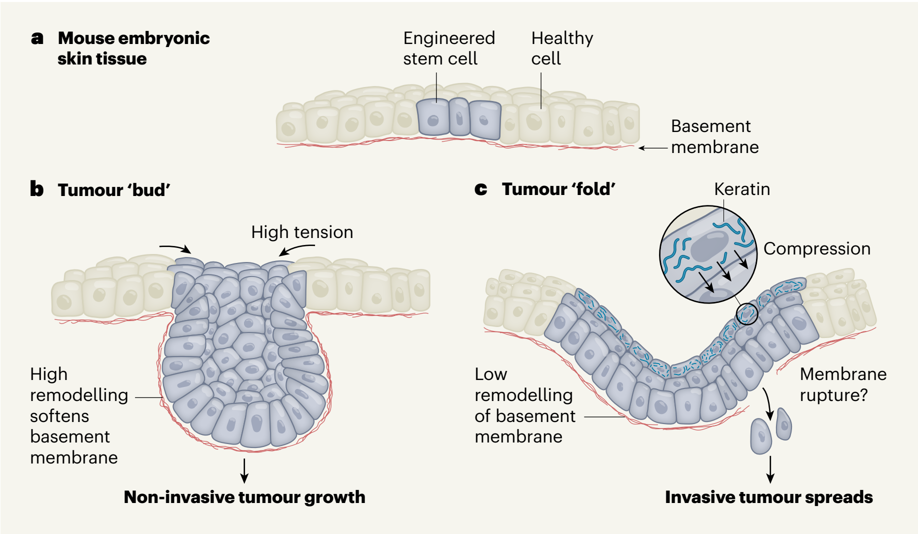


Figure 1 | Constructing the cellular architectures of cancer. a, Fiore *et al.*[1] engineered stem cells in embryonic mouse skin to have cancer-promoting mutations in the genes *SmoM2* or *HRas*. These mutant stem cells lie above a layer of extracellular matrix material called the basement membrane. **b**, The tumours in embryos with a *SmoM2* mutation resembled a benign, non-invasive cancer called basal cell carcinoma. These mutant cells actively produced and remodelled the basement membrane, rendering it elastic. The cancer cells generated forces as a result of cellular overcrowding, which buckled the basement membrane, creating a bud-shaped tumour, and produced tension at the boundary with the non-mutant cells. **c**, The tumours in embryos with an *HRas* mutation resembled a malignant, invasive cancer termed squamous cell carcinoma. These *HRas*-mutant cells produced less basement membrane than did the *SmoM2* mutant cells, and the membrane was rigid. The production of higher-than-normal levels of the protein keratin stiffened an upper layer of cells. Sandwiched between these two inflexible layers, the tumour could not easily dissipate the compressive forces exerted, producing an architecture of wave-like folds. The authors suggest that these forces might rupture the basement membrane, enabling invasion of the underlying tissue.

such boundary tension was not observed for SCC-like cells. Surprisingly, however, these differences were not decisive factors in driving tumour shape. The *HRas* mutation in the SCC-like tumours caused stiffening of the skin's outermost cellular layer by generating higher-than-normal levels of keratin proteins, a hallmark of this cancer. These keratin-rich cells were stiffer than the basal stem cells[4,5], and sandwiched the rapidly dividing SCC-like tumour cells between this stiff layer and the rigid basement membrane. The authors showed that both of these adjacent, rigid structures were needed to produce SCC architecture (Fig. 1).

The crucial role of mechanical force in generating biological structures has been highlighted in many contexts, including in the generation of various folds of epithelial cells. In particular, epithelial cells apply actin- and myosin-based contraction to engage in a tug-of-war against the underlying basement membrane. Depending on the mechanical properties of the surrounding structures and the amount of force generated by the epithelial cells, this results in either passive buckling or active folding of the epithelial tissue[6]. An exciting advance made with Fiore and colleagues' study is the merging of these models into a process that could be described as active buckling, in which cells exert specific contractile forces on their surroundings but also actively influence the mechanics of the underlying basement membrane to produce a specific tumour pattern.

The effect of mechanical forces on cancer has been addressed previously in other work. For example, in tubes formed of epithelial cells in the pancreas, tissue curvature is the key determinant that influences whether cancer grows inwards or outwards from such tubes[7]. One intriguing aspect of the work by Fiore *et al.* is their finding that a single cancer-promoting mutation suffices to orchestrate a stereotypical tumour architecture.

> **"The two cancer-promoting mutations had different effects on the production, turnover and stiffness of the basement membrane."**

Some questions remain to be answered. What are the signalling mechanisms responsible for changes in the production of basement membrane or the generation of a stiffness gradient in the multi-layered epidermis? Human tumours have complex mutational landscapes, so it will be interesting to assess what effect other genes that promote or hinder tumour development have on the processes that influence tumour shape.

Previous studies[8,9] of other systems provide clues about how physical changes can integrate with cellular signalling. During the development of chick feather follicle structures, mechanical compression triggers the movement of the protein β-catenin to the nucleus, where it drives a transcriptional response that enables cellular differentiation[8]. In mouse hair follicles, remodelling of the basement membrane modulates the Wnt and TGF-β signalling pathways needed to regulate stem-cell proliferation and subsequent tumour formation[9]. Thus, it is highly probable that, in cancers, mechanical forces are embedded within networks of biochemical signals, in which forces and signalling molecules might provide constant bidirectional feedback. It will be interesting to learn to what extent such feedback loops, if present, are similar in the context of normal tissue development and cancer.

The precise functional consequences of specific tumour architectures should be a key avenue for future research. Fiore and colleagues suggest that rupture of the basement membrane as a result of tissue forces, possibly accompanied by digestion of the extracellular matrix driven by protease enzymes, is responsible for the invasion of other tissues by SCC tumours. This fascinating hypothesis could have crucial implications if alterations in tumour architecture or basement membrane stiffness could provide early signs of invasion that might be used to predict the outcome of human cancers.

Karolina Punovuori and **Sara A. Wickström** are at the Helsinki Institute of Life Science and in the Stem Cells and Metabolism Research Program, Faculty of Medicine, University of Helsinki, Helsinki 00290, Finland, and also at the Wihuri Research Institute, Helsinki. **S.A.W.** is also at the Max Planck Institute for Biology of Ageing, Cologne, Germany.
e-mail: sara.wickstrom@helsinki.fi

1. Fiore, V. F. *et al. Nature* **585**, 433–439 (2020).
2. Gonzales, K. A. U. & Fuchs, E. *Dev. Cell* **43**, 387–401 (2017).
3. Sánchez-Danés, A. & Blainpain, C. *Nature Rev. Cancer* **18**, 549–561 (2018).
4. Miroshnikova, Y. A. *et al. Nature Cell Biol.* **20**, 69–80 (2018).
5. Seltmann, K., Fritsch, A. W., Käs, J. A. & Magin. T. M. *Proc. Natl Acad. Sci. USA* **110**, 18507–18512 (2013).
6. Mao, Y. & Baum, B. *Dev. Biol.* **401**, 92–102 (2015).
7. Messal, H. A. *et al. Nature* **566**, 126–132 (2019).
8. Shyer, A. E. *et al. Science* **357**, 811–815 (2017).
9. Morgner, J. *et al. Nature Commun.* **6**, 8198 (2015).

Array programming with NumPy

https://doi.org/10.1038/s41586-020-2649-2

Received: 21 February 2020

Accepted: 17 June 2020

Charles R. Harris[1], K. Jarrod Millman[2,3,4] ✉, Stéfan J. van der Walt[2,4,5] ✉, Ralf Gommers[6] ✉, Pauli Virtanen[7,8], David Cournapeau[9], Eric Wieser[10], Julian Taylor[11], Sebastian Berg[4], Nathaniel J. Smith[12], Robert Kern[13], Matti Picus[4], Stephan Hoyer[14], Marten H. van Kerkwijk[15], Matthew Brett[2,16], Allan Haldane[17], Jaime Fernández del Río[18], Mark Wiebe[19,20], Pearu Peterson[6,21,22], Pierre Gérard-Marchant[23,24], Kevin Sheppard[25], Tyler Reddy[26], Warren Weckesser[4], Hameer Abbasi[6], Christoph Gohlke[27] & Travis E. Oliphant[6]

Array programming provides a powerful, compact and expressive syntax for accessing, manipulating and operating on data in vectors, matrices and higher-dimensional arrays. NumPy is the primary array programming library for the Python language. It has an essential role in research analysis pipelines in fields as diverse as physics, chemistry, astronomy, geoscience, biology, psychology, materials science, engineering, finance and economics. For example, in astronomy, NumPy was an important part of the software stack used in the discovery of gravitational waves[1] and in the first imaging of a black hole[2]. Here we review how a few fundamental array concepts lead to a simple and powerful programming paradigm for organizing, exploring and analysing scientific data. NumPy is the foundation upon which the scientific Python ecosystem is constructed. It is so pervasive that several projects, targeting audiences with specialized needs, have developed their own NumPy-like interfaces and array objects. Owing to its central position in the ecosystem, NumPy increasingly acts as an interoperability layer between such array computation libraries and, together with its application programming interface (API), provides a flexible framework to support the next decade of scientific and industrial analysis.

Two Python array packages existed before NumPy. The Numeric package was developed in the mid-1990s and provided array objects and array-aware functions in Python. It was written in C and linked to standard fast implementations of linear algebra[3,4]. One of its earliest uses was to steer C++ applications for inertial confinement fusion research at Lawrence Livermore National Laboratory[5]. To handle large astronomical images coming from the Hubble Space Telescope, a reimplementation of Numeric, called Numarray, added support for structured arrays, flexible indexing, memory mapping, byte-order variants, more efficient memory use, flexible IEEE 754-standard error-handling capabilities, and better type-casting rules[6]. Although Numarray was highly compatible with Numeric, the two packages had enough differences that it divided the community; however, in 2005 NumPy emerged as a 'best of both worlds' unification[7]—combining the features of Numarray with the small-array performance of Numeric and its rich C API.

Now, 15 years later, NumPy underpins almost every Python library that does scientific or numerical computation[8–11], including SciPy[12], Matplotlib[13], pandas[14], scikit-learn[15] and scikit-image[16]. NumPy is a community-developed, open-source library, which provides a multidimensional Python array object along with array-aware functions that operate on it. Because of its inherent simplicity, the NumPy array is the de facto exchange format for array data in Python.

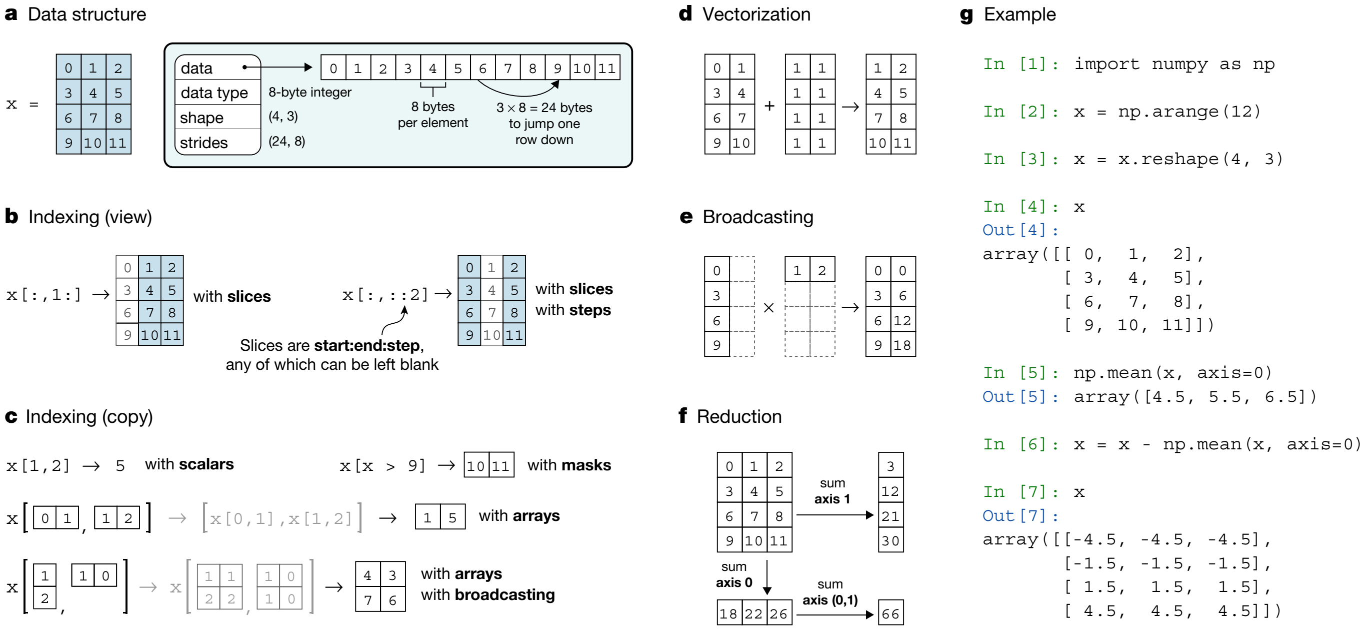
NumPy operates on in-memory arrays using the central processing unit (CPU). To utilize modern, specialized storage and hardware, there has been a recent proliferation of Python array packages. Unlike with the Numarray–Numeric divide, it is now much harder for these new libraries to fracture the user community—given how much work is already built on top of NumPy. However, to provide the community with access to new and exploratory technologies, NumPy is transitioning into a central coordinating mechanism that specifies a well defined array programming API and dispatches it, as appropriate, to specialized array implementations.

NumPy arrays

The NumPy array is a data structure that efficiently stores and accesses multidimensional arrays[17] (also known as tensors), and enables a wide variety of scientific computation. It consists of a pointer to memory, along with metadata used to interpret the data stored there, notably 'data type', 'shape' and 'strides' (Fig. 1a).

---

[1]Independent researcher, Logan, UT, USA. [2]Brain Imaging Center, University of California, Berkeley, Berkeley, CA, USA. [3]Division of Biostatistics, University of California, Berkeley, Berkeley, CA, USA. [4]Berkeley Institute for Data Science, University of California, Berkeley, Berkeley, CA, USA. [5]Applied Mathematics, Stellenbosch University, Stellenbosch, South Africa. [6]Quansight, Austin, TX, USA. [7]Department of Physics, University of Jyväskylä, Jyväskylä, Finland. [8]Nanoscience Center, University of Jyväskylä, Jyväskylä, Finland. [9]Mercari JP, Tokyo, Japan. [10]Department of Engineering, University of Cambridge, Cambridge, UK. [11]Independent researcher, Karlsruhe, Germany. [12]Independent researcher, Berkeley, CA, USA. [13]Enthought, Austin, TX, USA. [14]Google Research, Mountain View, CA, USA. [15]Department of Astronomy and Astrophysics, University of Toronto, Toronto, Ontario, Canada. [16]School of Psychology, University of Birmingham, Edgbaston, Birmingham, UK. [17]Department of Physics, Temple University, Philadelphia, PA, USA. [18]Google, Zurich, Switzerland. [19]Department of Physics and Astronomy, The University of British Columbia, Vancouver, British Columbia, Canada. [20]Amazon, Seattle, WA, USA. [21]Independent researcher, Saue, Estonia. [22]Department of Mechanics and Applied Mathematics, Institute of Cybernetics at Tallinn Technical University, Tallinn, Estonia. [23]Department of Biological and Agricultural Engineering, University of Georgia, Athens, GA, USA. [24]France-IX Services, Paris, France. [25]Department of Economics, University of Oxford, Oxford, UK. [26]CCS-7, Los Alamos National Laboratory, Los Alamos, NM, USA. [27]Laboratory for Fluorescence Dynamics, Biomedical Engineering Department, University of California, Irvine, Irvine, CA, USA. ✉e-mail: millman@berkeley.edu; stefanv@berkeley.edu; ralf.gommers@gmail.com



```
In [1]: import numpy as np

In [2]: x = np.arange(12)

In [3]: x = x.reshape(4, 3)

In [4]: x
Out[4]:
array([[ 0,  1,  2],
       [ 3,  4,  5],
       [ 6,  7,  8],
       [ 9, 10, 11]])

In [5]: np.mean(x, axis=0)
Out[5]: array([4.5, 5.5, 6.5])

In [6]: x = x - np.mean(x, axis=0)

In [7]: x
Out[7]:
array([[-4.5, -4.5, -4.5],
       [-1.5, -1.5, -1.5],
       [ 1.5,  1.5,  1.5],
       [ 4.5,  4.5,  4.5]])
```

Fig. 1 | The NumPy array incorporates several fundamental array concepts. **a**, The NumPy array data structure and its associated metadata fields. **b**, Indexing an array with slices and steps. These operations return a 'view' of the original data. **c**, Indexing an array with masks, scalar coordinates or other arrays, so that it returns a 'copy' of the original data. In the bottom example, an array is indexed with other arrays; this broadcasts the indexing arguments before performing the lookup. **d**, Vectorization efficiently applies operations to groups of elements. **e**, Broadcasting in the multiplication of two-dimensional arrays. **f**, Reduction operations act along one or more axes. In this example, an array is summed along select axes to produce a vector, or along two axes consecutively to produce a scalar. **g**, Example NumPy code, illustrating some of these concepts.

The data type describes the nature of elements stored in an array. An array has a single data type, and each element of an array occupies the same number of bytes in memory. Examples of data types include real and complex numbers (of lower and higher precision), strings, timestamps and pointers to Python objects.

The shape of an array determines the number of elements along each axis, and the number of axes is the dimensionality of the array. For example, a vector of numbers can be stored as a one-dimensional array of shape N, whereas colour videos are four-dimensional arrays of shape $(T, M, N, 3)$.

Strides are necessary to interpret computer memory, which stores elements linearly, as multidimensional arrays. They describe the number of bytes to move forward in memory to jump from row to row, column to column, and so forth. Consider, for example, a two-dimensional array of floating-point numbers with shape (4, 3), where each element occupies 8 bytes in memory. To move between consecutive columns, we need to jump forward 8 bytes in memory, and to access the next row, 3 × 8 = 24 bytes. The strides of that array are therefore (24, 8). NumPy can store arrays in either C or Fortran memory order, iterating first over either rows or columns. This allows external libraries written in those languages to access NumPy array data in memory directly.

Users interact with NumPy arrays using 'indexing' (to access subarrays or individual elements), 'operators' (for example, +, − and × for vectorized operations and @ for matrix multiplication), as well as 'array-aware functions'; together, these provide an easily readable, expressive, high-level API for array programming while NumPy deals with the underlying mechanics of making operations fast.

Indexing an array returns single elements, subarrays or elements that satisfy a specific condition (Fig. 1b). Arrays can even be indexed using other arrays (Fig. 1c). Wherever possible, indexing that retrieves a subarray returns a 'view' on the original array such that data are shared between the two arrays. This provides a powerful way to operate on subsets of array data while limiting memory usage.

To complement the array syntax, NumPy includes functions that perform vectorized calculations on arrays, including arithmetic, statistics and trigonometry (Fig. 1d). Vectorization—operating on entire arrays rather than their individual elements—is essential to array programming. This means that operations that would take many tens of lines to express in languages such as C can often be implemented as a single, clear Python expression. This results in concise code and frees users to focus on the details of their analysis, while NumPy handles looping over array elements near-optimally—for example, taking strides into consideration to best utilize the computer's fast cache memory.

When performing a vectorized operation (such as addition) on two arrays with the same shape, it is clear what should happen. Through 'broadcasting' NumPy allows the dimensions to differ, and produces results that appeal to intuition. A trivial example is the addition of a scalar value to an array, but broadcasting also generalizes to more complex examples such as scaling each column of an array or generating a grid of coordinates. In broadcasting, one or both arrays are virtually duplicated (that is, without copying any data in memory), so that the shapes of the operands match (Fig. 1d). Broadcasting is also applied when an array is indexed using arrays of indices (Fig. 1c).

Other array-aware functions, such as sum, mean and maximum, perform element-by-element 'reductions', aggregating results across one, multiple or all axes of a single array. For example, summing an n-dimensional array over d axes results in an array of dimension $n - d$ (Fig. 1f).

NumPy also includes array-aware functions for creating, reshaping, concatenating and padding arrays; searching, sorting and counting data; and reading and writing files. It provides extensive support for generating pseudorandom numbers, includes an assortment of probability distributions, and performs accelerated linear algebra, using one of several backends such as OpenBLAS[18,19] or Intel MKL optimized for the CPUs at hand (see Supplementary Methods for more details).

Altogether, the combination of a simple in-memory array representation, a syntax that closely mimics mathematics, and a variety of array-aware utility functions forms a productive and powerfully expressive array programming language.

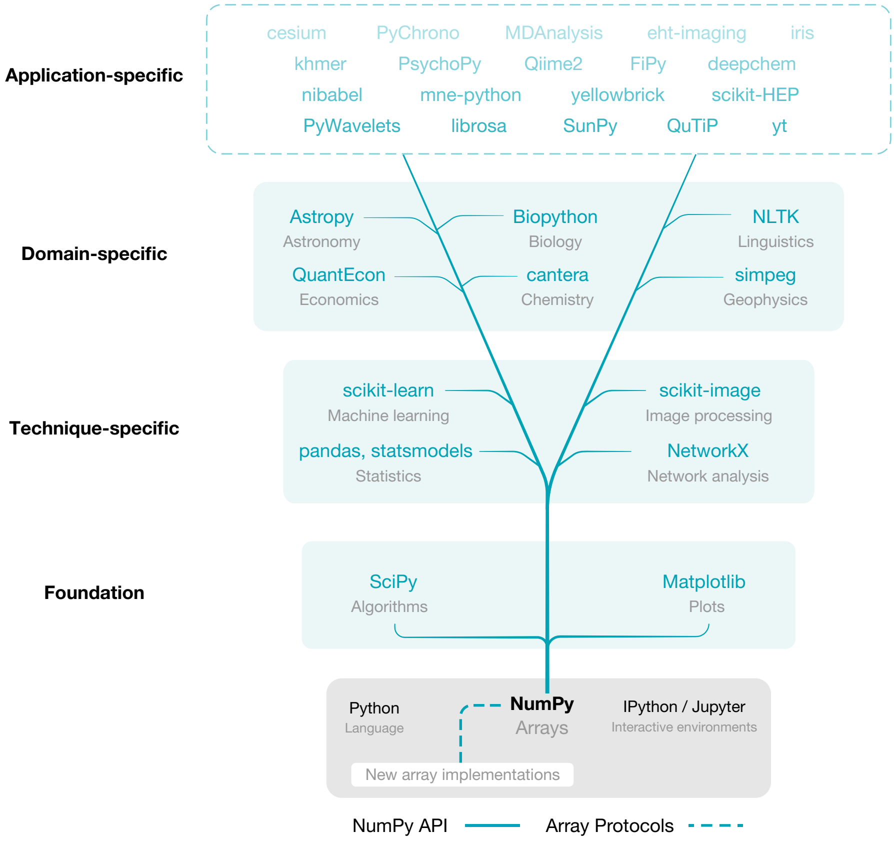


Fig. 2 | NumPy is the base of the scientific Python ecosystem. Essential libraries and projects that depend on NumPy's API gain access to new array implementations that support NumPy's array protocols (Fig. 3).

Scientific Python ecosystem

Python is an open-source, general-purpose interpreted programming language well suited to standard programming tasks such as cleaning data, interacting with web resources and parsing text. Adding fast array operations and linear algebra enables scientists to do all their work within a single programming language—one that has the advantage of being famously easy to learn and teach, as witnessed by its adoption as a primary learning language in many universities.

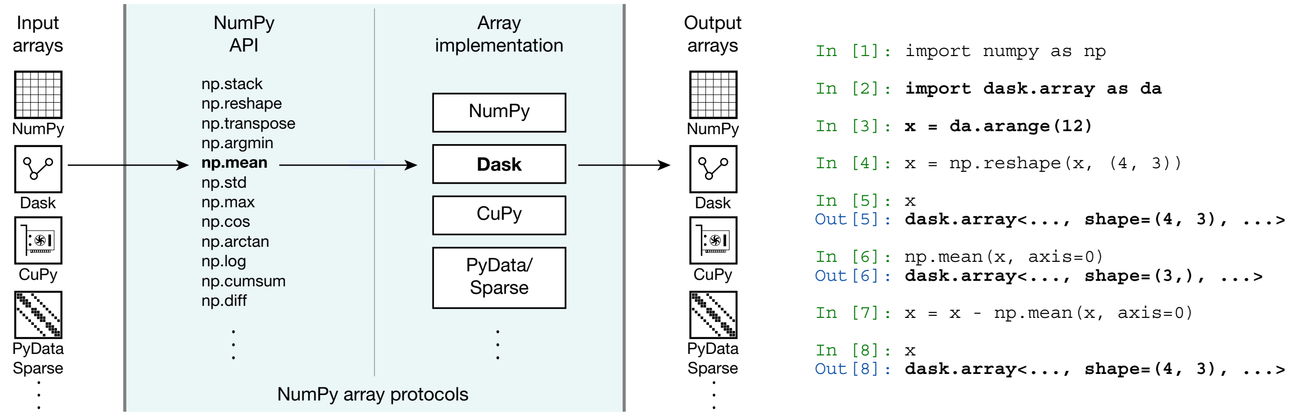
Even though NumPy is not part of Python's standard library, it benefits from a good relationship with the Python developers. Over the years, the Python language has added new features and special syntax so that NumPy would have a more succinct and easier-to-read array notation. However, because it is not part of the standard library, NumPy is able to dictate its own release policies and development patterns.

SciPy and Matplotlib are tightly coupled with NumPy in terms of history, development and use. SciPy provides fundamental algorithms for scientific computing, including mathematical, scientific and engineering routines. Matplotlib generates publication-ready figures and visualizations. The combination of NumPy, SciPy and Matplotlib, together with an advanced interactive environment such as IPython[20] or Jupyter[21], provides a solid foundation for array programming in Python. The scientific Python ecosystem (Fig. 2) builds on top of this foundation to provide several, widely used technique-specific libraries[15,16,22], that in turn underlie numerous domain-specific projects[23–28]. NumPy, at the base of the ecosystem of array-aware libraries, sets documentation standards, provides array testing infrastructure and adds build support for Fortran and other compilers.

Many research groups have designed large, complex scientific libraries that add application-specific functionality to the ecosystem. For example, the eht-imaging library[29], developed by the Event Horizon Telescope collaboration for radio interferometry imaging, analysis and simulation, relies on many lower-level components of the scientific Python ecosystem. In particular, the EHT collaboration used this library for the first imaging of a black hole. Within eht-imaging, NumPy arrays are used to store and manipulate numerical data at every step in the processing chain: from raw data through calibration and image reconstruction. SciPy supplies tools for general image-processing tasks such as filtering and image alignment, and scikit-image, an image-processing library that extends SciPy, provides higher-level functionality such as edge filters and Hough transforms. The 'scipy.optimize' module performs mathematical optimization. NetworkX[22], a package for complex network analysis, is used to verify image comparison consistency. Astropy[23,24] handles standard astronomical file formats and computes time–coordinate transformations. Matplotlib is used to visualize data and to generate the final image of the black hole.

The interactive environment created by the array programming foundation and the surrounding ecosystem of tools—inside of IPython or Jupyter—is ideally suited to exploratory data analysis. Users can fluidly inspect, manipulate and visualize their data, and rapidly iterate to refine programming statements. These statements are then stitched together into imperative or functional programs, or notebooks containing both computation and narrative. Scientific computing beyond exploratory work is often done in a text editor or an integrated development environment (IDE) such as Spyder. This rich and productive environment has made Python popular for scientific research.

To complement this facility for exploratory work and rapid prototyping, NumPy has developed a culture of using time-tested software engineering practices to improve collaboration and reduce error[30]. This culture is not only adopted by leaders in the project but also enthusiastically taught to newcomers. The NumPy team was early to adopt distributed revision control and code review to improve collaboration



```
In [1]: import numpy as np

In [2]: import dask.array as da

In [3]: x = da.arange(12)

In [4]: x = np.reshape(x, (4, 3))

In [5]: x
Out[5]: dask.array<..., shape=(4, 3), ...>

In [6]: np.mean(x, axis=0)
Out[6]: dask.array<..., shape=(3,), ...>

In [7]: x = x - np.mean(x, axis=0)

In [8]: x
Out[8]: dask.array<..., shape=(4, 3), ...>
```

Fig. 3 | NumPy's API and array protocols expose new arrays to the ecosystem. In this example, NumPy's 'mean' function is called on a Dask array. The call succeeds by dispatching to the appropriate library implementation (in this case, Dask) and results in a new Dask array. Compare this code to the example code in Fig. 1g.

on code, and continuous testing that runs an extensive battery of automated tests for every proposed change to NumPy. The project also has comprehensive, high-quality documentation, integrated with the source code[31–33].

This culture of using best practices for producing reliable scientific software has been adopted by the ecosystem of libraries that build on NumPy. For example, in a recent award given by the Royal Astronomical Society to Astropy, they state: "The Astropy Project has provided hundreds of junior scientists with experience in professional-standard software development practices including use of version control, unit testing, code review and issue tracking procedures. This is a vital skill set for modern researchers that is often missing from formal university education in physics or astronomy"[34]. Community members explicitly work to address this lack of formal education through courses and workshops[35–37].

The recent rapid growth of data science, machine learning and artificial intelligence has further and dramatically boosted the scientific use of Python. Examples of its important applications, such as the eht-imaging library, now exist in almost every discipline in the natural and social sciences. These tools have become the primary software environment in many fields. NumPy and its ecosystem are commonly taught in university courses, boot camps and summer schools, and are the focus of community conferences and workshops worldwide. NumPy and its API have become truly ubiquitous.

Array proliferation and interoperability

NumPy provides in-memory, multidimensional, homogeneously typed (that is, single-pointer and strided) arrays on CPUs. It runs on machines ranging from embedded devices to the world's largest supercomputers, with performance approaching that of compiled languages. For most its existence, NumPy addressed the vast majority of array computation use cases.

However, scientific datasets now routinely exceed the memory capacity of a single machine and may be stored on multiple machines or in the cloud. In addition, the recent need to accelerate deep-learning and artificial intelligence applications has led to the emergence of specialized accelerator hardware, including graphics processing units (GPUs), tensor processing units (TPUs) and field-programmable gate arrays (FPGAs). Owing to its in-memory data model, NumPy is currently unable to directly utilize such storage and specialized hardware. However, both distributed data and also the parallel execution of GPUs, TPUs and FPGAs map well to the paradigm of array programming: therefore leading to a gap between available modern hardware architectures and the tools necessary to leverage their computational power.

The community's efforts to fill this gap led to a proliferation of new array implementations. For example, each deep-learning framework created its own arrays; the PyTorch[38], Tensorflow[39], Apache MXNet[40] and JAX arrays all have the capability to run on CPUs and GPUs in a distributed fashion, using lazy evaluation to allow for additional performance optimizations. SciPy and PyData/Sparse both provide sparse arrays, which typically contain few non-zero values and store only those in memory for efficiency. In addition, there are projects that build on NumPy arrays as data containers, and extend its capabilities. Distributed arrays are made possible that way by Dask, and labelled arrays—referring to dimensions of an array by name rather than by index for clarity, compare x[:, 1] versus x.loc[:, 'time']—by xarray[41].

Such libraries often mimic the NumPy API, because this lowers the barrier to entry for newcomers and provides the wider community with a stable array programming interface. This, in turn, prevents disruptive schisms such as the divergence between Numeric and Numarray. But exploring new ways of working with arrays is experimental by nature and, in fact, several promising libraries (such as Theano and Caffe) have already ceased development. And each time that a user decides to try a new technology, they must change import statements and ensure that the new library implements all the parts of the NumPy API they currently use.

Ideally, operating on specialized arrays using NumPy functions or semantics would simply work, so that users could write code once, and would then benefit from switching between NumPy arrays, GPU arrays, distributed arrays and so forth as appropriate. To support array operations between external array objects, NumPy therefore added the capability to act as a central coordination mechanism with a well specified API (Fig. 2).

To facilitate this interoperability, NumPy provides 'protocols' (or contracts of operation), that allow for specialized arrays to be passed to NumPy functions (Fig. 3). NumPy, in turn, dispatches operations to the originating library, as required. Over four hundred of the most popular NumPy functions are supported. The protocols are implemented by widely used libraries such as Dask, CuPy, xarray and PyData/Sparse. Thanks to these developments, users can now, for example, scale their computation from a single machine to distributed systems using Dask. The protocols also compose well, allowing users to redeploy NumPy code at scale on distributed, multi-GPU systems via, for instance, CuPy arrays embedded in Dask arrays. Using NumPy's high-level API, users can leverage highly parallel code execution on multiple systems with millions of cores, all with minimal code changes[42].

These array protocols are now a key feature of NumPy, and are expected to only increase in importance. The NumPy developers—many of whom are authors of this Review—iteratively refine and add protocol designs to improve utility and simplify adoption.

Discussion

NumPy combines the expressive power of array programming, the performance of C, and the readability, usability and versatility of Python in a mature, well tested, well documented and community-developed library. Libraries in the scientific Python ecosystem provide fast implementations of most important algorithms. Where extreme optimization is warranted, compiled languages can be used, such as Cython[43], Numba[44] and Pythran[45]; these languages extend Python and transparently accelerate bottlenecks. Owing to NumPy's simple memory model, it is easy to write low-level, hand-optimized code, usually in C or Fortran, to manipulate NumPy arrays and pass them back to Python. Furthermore, using array protocols, it is possible to utilize the full spectrum of specialized hardware acceleration with minimal changes to existing code.

NumPy was initially developed by students, faculty and researchers to provide an advanced, open-source array programming library for Python, which was free to use and unencumbered by license servers and software protection dongles. There was a sense of building something consequential together for the benefit of many others. Participating in such an endeavour, within a welcoming community of like-minded individuals, held a powerful attraction for many early contributors.

These user–developers frequently had to write code from scratch to solve their own or their colleagues' problems—often in low-level languages that preceded Python, such as Fortran[46] and C. To them, the advantages of an interactive, high-level array library were evident. The design of this new tool was informed by other powerful interactive programming languages for scientific computing such as Basis[47–50], Yorick[51], R[52] and APL[53], as well as commercial languages and environments such as IDL (Interactive Data Language) and MATLAB.

What began as an attempt to add an array object to Python became the foundation of a vibrant ecosystem of tools. Now, a large amount of scientific work depends on NumPy being correct, fast and stable. It is no longer a small community project, but core scientific infrastructure.

The developer culture has matured: although initial development was highly informal, NumPy now has a roadmap and a process for proposing and discussing large changes. The project has formal governance structures and is fiscally sponsored by NumFOCUS, a nonprofit that promotes open practices in research, data and scientific computing. Over the past few years, the project attracted its first funded development, sponsored by the Moore and Sloan Foundations, and received an award as part of the Chan Zuckerberg Initiative's Essentials of Open Source Software programme. With this funding, the project was (and is) able to have sustained focus over multiple months to implement substantial new features and improvements. That said, the development of NumPy still depends heavily on contributions made by graduate students and researchers in their free time (see Supplementary Methods for more details).

NumPy is no longer merely the foundational array library underlying the scientific Python ecosystem, but it has become the standard API for tensor computation and a central coordinating mechanism between array types and technologies in Python. Work continues to expand on and improve these interoperability features.

Over the next decade, NumPy developers will face several challenges. New devices will be developed, and existing specialized hardware will evolve to meet diminishing returns on Moore's law. There will be more, and a wider variety of, data science practitioners, a large proportion of whom will use NumPy. The scale of scientific data gathering will continue to increase, with the adoption of devices and instruments such as light-sheet microscopes and the Large Synoptic Survey Telescope (LSST)[54]. New generation languages, interpreters and compilers, such as Rust[55], Julia[56] and LLVM[57], will create new concepts and data structures, and determine their viability.

Through the mechanisms described in this Review, NumPy is poised to embrace such a changing landscape, and to continue playing a leading part in interactive scientific computation, although to do so will require sustained funding from government, academia and industry. But, importantly, for NumPy to meet the needs of the next decade of data science, it will also need a new generation of graduate students and community contributors to drive it forward.

1. Abbott, B. P. et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett*. **116**, 061102 (2016).
2. Chael, A. et al. High-resolution linear polarimetric imaging for the Event Horizon Telescope. *Astrophys. J*. **286**, 11 (2016).
3. Dubois, P. F., Hinsen, K. & Hugunin, J. Numerical Python. *Comput. Phys*. **10**, 262–267 (1996).
4. Ascher, D., Dubois, P. F., Hinsen, K., Hugunin, J. & Oliphant, T. E. *An Open Source Project: Numerical Python* (Lawrence Livermore National Laboratory, 2001).
5. Yang, T.-Y., Furnish, G. & Dubois, P. F. Steering object-oriented scientific computations. In *Proc. TOOLS USA 97. Intl Conf. Technology of Object Oriented Systems and Languages* (eds Ege, R., Singh, M. & Meyer, B.) 112–119 (IEEE, 1997).
6. Greenfield, P., Miller, J. T., Hsu, J. & White, R. L. numarray: a new scientific array package for Python. In *PyCon DC 2003* http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.112.9899 (2003).
7. Oliphant, T. E. *Guide to NumPy* 1st edn (Trelgol Publishing, 2006).
8. Dubois, P. F. Python: batteries included. *Comput. Sci. Eng*. **9**, 7–9 (2007).
9. Oliphant, T. E. Python for scientific computing. *Comput. Sci. Eng*. **9**, 10–20 (2007).
10. Millman, K. J. & Aivazis, M. Python for scientists and engineers. *Comput. Sci. Eng*. **13**, 9–12 (2011).
11. Pérez, F., Granger, B. E. & Hunter, J. D. Python: an ecosystem for scientific computing. *Comput. Sci. Eng*. **13**, 13–21 (2011).
Explains why the scientific Python ecosystem is a highly productive environment for research.
12. Virtanen, P. et al. SciPy 1.0—fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020); correction **17**, 352 (2020).
Introduces the SciPy library and includes a more detailed history of NumPy and SciPy.
13. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng*. **9**, 90–95 (2007).
14. McKinney, W. Data structures for statistical computing in Python. In *Proc. 9th Python in Science Conf*. (eds van der Walt, S. & Millman, K. J.) 56–61 (2010).
15. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res*. **12**, 2825–2830 (2011).
16. van der Walt, S. et al. scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).
17. van der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng*. **13**, 22–30 (2011).
Discusses the NumPy array data structure with a focus on how it enables efficient computation.
18. Wang, Q., Zhang, X., Zhang, Y. & Yi, Q. AUGEM: automatically generate high performance dense linear algebra kernels on x86 CPUs. In *SC'13: Proc. Intl Conf. High Performance Computing, Networking, Storage and Analysis* 25 (IEEE, 2013).
19. Xianyi, Z., Qian, W. & Yunquan, Z. Model-driven level 3 BLAS performance optimization on Loongson 3A processor. In *2012 IEEE 18th Intl Conf. Parallel and Distributed Systems* 684–691 (IEEE, 2012).
20. Pérez, F. & Granger, B. E. IPython: a system for interactive scientific computing. *Comput. Sci. Eng*. **9**, 21–29 (2007).
21. Kluyver, T. et al. Jupyter Notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (eds Loizides, F. & Schmidt, B.) 87–90 (IOS Press, 2016).
22. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. In *Proc. 7th Python in Science Conf*. (eds Varoquaux, G., Vaught, T. & Millman, K. J.) 11–15 (2008).
23. Astropy Collaboration et al. Astropy: a community Python package for astronomy. *Astron. Astrophys*. **558**, A33 (2013).
24. Price-Whelan, A. M. et al. The Astropy Project: building an open-science project and status of the v2.0 core package. *Astron. J*. **156**, 123 (2018).
25. Cock, P. J. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
26. Millman, K. J. & Brett, M. Analysis of functional magnetic resonance imaging in Python. *Comput. Sci. Eng*. **9**, 52–55 (2007).
27. The SunPy Community et al. SunPy—Python for solar physics. *Comput. Sci. Discov*. **8**, 014009 (2015).
28. Hamman, J., Rocklin, M. & Abernathy, R. Pangeo: a big-data ecosystem for scalable Earth system science. In *EGU General Assembly Conf. Abstracts* 12146 (2018).
29. Chael, A. A. et al. ehtim: imaging, analysis, and simulation software for radio interferometry. *Astrophysics Source Code Library* https://ascl.net/1904.004 (2019).
30. Millman, K. J. & Pérez, F. Developing open source scientific practice. In *Implementing Reproducible Research* (eds Stodden, V., Leisch, F. & Peng, R. D.) 149–183 (CRC Press, 2014).
Describes the software engineering practices embraced by the NumPy and SciPy communities with a focus on how these practices improve research.
31. van der Walt, S. The SciPy Documentation Project (technical overview). In *Proc. 7th Python in Science Conf. (SciPy 2008)* (eds Varoquaux, G., Vaught, T. & Millman, K. J.) 27–28 (2008).
32. Harrington, J. The SciPy Documentation Project. In *Proc. 7th Python in Science Conference (SciPy 2008)* (eds Varoquaux, G., Vaught, T. & Millman, K. J.) 33–35 (2008).
33. Harrington, J. & Goldsmith, D. Progress report: NumPy and SciPy documentation in 2009. In *Proc. 8th Python in Science Conf. (SciPy 2009)* (eds Varoquaux, G., van der Walt, S. & Millman, K. J.) 84–87 (2009).
34. Royal Astronomical Society *Report of the RAS 'A' Awards Committee 2020: Astropy Project: 2020 Group Achievement Award (A)* https://ras.ac.uk/sites/default/files/2020-01/Group%20Award%20-%20Astropy.pdf (2020).
35. Wilson, G. Software carpentry: getting scientists to write better code by making them more productive. *Comput. Sci. Eng*. **8**, 66–69 (2006).

36. Hannay, J. E. et al. How do scientists develop and use scientific software? In *Proc. 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering* 1–8 (IEEE, 2009).
37. Millman, K. J., Brett, M., Barnowski, R. & Poline, J.-B. Teaching computational reproducibility for neuroimaging. *Front. Neurosci.* **12**, 727 (2018).
38. Paszke, A. et al. Pytorch: an imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32* (eds Wallach, H. et al.) 8024–8035 (Neural Information Processing Systems, 2019).
39. Abadi, M. et al. TensorFlow: a system for large-scale machine learning. In *OSDI'16: Proc. 12th USENIX Conf. Operating Systems Design and Implementation* (chairs Keeton, K. & Roscoe, T.) 265–283 (USENIX Association, 2016).
40. Chen, T. et al. MXNet: a flexible and efficient machine learning library for heterogeneous distributed systems. Preprint at http://www.arxiv.org/abs/1512.01274 (2015).
41. Hoyer, S. & Hamman, J. xarray: N–D labeled arrays and datasets in Python. *J. Open Res. Softw*. **5**, 10 (2017).
42. Entschev, P. Distributed multi-GPU computing with Dask, CuPy and RAPIDS. In *EuroPython 2019* https://ep2019.europython.eu/media/conference/slides/fX8dJsD-distributed-multi-gpu-computing-with-dask-cupy-and-rapids.pdf (2019).
43. Behnel, S. et al. Cython: the best of both worlds. *Comput. Sci. Eng*. **13**, 31–39 (2011).
44. Lam, S. K., Pitrou, A. & Seibert, S. Numba: a LLVM-based Python JIT compiler. In *Proc. Second Workshop on the LLVM Compiler Infrastructure in HPC, LLVM '15* 7:1–7:6 (ACM, 2015).
45. Guelton, S. et al. Pythran: enabling static optimization of scientific Python programs. *Comput. Sci. Discov*. **8**, 014001 (2015).
46. Dongarra, J., Golub, G. H., Grosse, E., Moler, C. & Moore, K. Netlib and NA-Net: building a scientific computing community. *IEEE Ann. Hist. Comput*. **30**, 30–41 (2008).
47. Barrett, K. A., Chiu, Y. H., Painter, J. F., Motteler, Z. C. & Dubois, P. F. *Basis System, Part I: Running a Basis Program—A Tutorial for Beginners* UCRL-MA-118543, Vol. 1 (Lawrence Livermore National Laboratory 1995).
48. Dubois, P. F. & Motteler, Z. *Basis System, Part II: Basis Language Reference Manual* UCRL-MA-118543, Vol. 2 (Lawrence Livermore National Laboratory, 1995).
49. Chiu, Y. H. & Dubois, P. F. *Basis System, Part III: EZN User Manual* UCRL-MA-118543, Vol. 3 (Lawrence Livermore National Laboratory, 1995).
50. Chiu, Y. H. & Dubois, P. F. *Basis System, Part IV: EZD User Manual* UCRL-MA-118543, Vol. 4 (Lawrence Livermore National Laboratory, 1995).
51. Munro, D. H. & Dubois, P. F. Using the Yorick interpreted language. *Comput. Phys*. **9**, 609–615 (1995).
52. Ihaka, R. & Gentleman, R. R: a language for data analysis and graphics. *J. Comput. Graph. Stat*. **5**, 299–314 (1996).
53. Iverson, K. E. A programming language. In *Proc. 1962 Spring Joint Computer Conf*. 345–351 (1962).
54. Jenness, T. et al. LSST data management software development practices and tools. In *Proc. SPIE 10707, Software and Cyberinfrastructure for Astronomy V* 1070709 (SPIE and International Society for Optics and Photonics, 2018).
55. Matsakis, N. D. & Klock, F. S. The Rust language. *Ada Letters* **34**, 103–104 (2014).
56. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: a fresh approach to numerical computing. *SIAM Rev*. **59**, 65–98 (2017).
57. Lattner, C. & Adve, V. LLVM: a compilation framework for lifelong program analysis and transformation. In *Proc. 2004 Intl Symp. Code Generation and Optimization (CGO'04)* 75–88 (IEEE, 2004).

Acknowledgements We thank R. Barnowski, P. Dubois, M. Eickenberg, and P. Greenfield, who suggested text and provided helpful feedback on the manuscript. K.J.M. and S.J.v.d.W. were funded in part by the Gordon and Betty Moore Foundation through grant GBMF3834 and by the Alfred P. Sloan Foundation through grant 2013-10-27 to the University of California, Berkeley. S.J.v.d.W., S.B., M.P. and W.W. were funded in part by the Gordon and Betty Moore Foundation through grant GBMF5447 and by the Alfred P. Sloan Foundation through grant G-2017-9960 to the University of California, Berkeley.

Author contributions K.J.M. and S.J.v.d.W. composed the manuscript with input from others. S.B., R.G., K.S., W.W., M.B. and T.R. contributed text. All authors contributed substantial code, documentation and/or expertise to the NumPy project. All authors reviewed the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2649-2.
Correspondence and requests for materials should be addressed to K.J.M., S.J.v.W. or R.G.
Peer review information *Nature* thanks Edouard Duchesnay, Alan Edelman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.

A giant planet candidate transiting a white dwarf

https://doi.org/10.1038/s41586-020-2713-y

Received: 16 March 2020

Accepted: 15 July 2020

Andrew Vanderburg[1,2]✉, Saul A. Rappaport[3], Siyi Xu[4], Ian J. M. Crossfield[5], Juliette C. Becker[6], Bruce Gary[7], Felipe Murgas[8,9], Simon Blouin[10], Thomas G. Kaye[11,12], Enric Palle[8,9], Carl Melis[13], Brett M. Morris[14], Laura Kreidberg[15,16], Varoujan Gorjian[17], Caroline V. Morley[2], Andrew W. Mann[18], Hannu Parviainen[8,9], Logan A. Pearce[19], Elisabeth R. Newton[20], Andreia Carrillo[2], Ben Zuckerman[21], Lorne Nelson[22], Greg Zeimann[23], Warren R. Brown[16], René Tronsgaard[24], Beth Klein[21], George R. Ricker[3], Roland K. Vanderspek[3], David W. Latham[16], Sara Seager[3,25,26], Joshua N. Winn[27], Jon M. Jenkins[28], Fred C. Adams[29,30], Björn Benneke[31,32], David Berardo[3], Lars A. Buchhave[24], Douglas A. Caldwell[28,33], Jessie L. Christiansen[34], Karen A. Collins[16], Knicole D. Colón[35], Tansu Daylan[3], John Doty[36], Alexandra E. Doyle[37], Diana Dragomir[38], Courtney Dressing[39], Patrick Dufour[31,32], Akihiko Fukui[8,40], Ana Glidden[3,25], Natalia M. Guerrero[3], Xueying Guo[3], Kevin Heng[14], Andreea I. Henriksen[24], Chelsea X. Huang[3], Lisa Kaltenegger[41,42], Stephen R. Kane[43], John A. Lewis[16], Jack J. Lissauer[28], Farisa Morales[17,44], Norio Narita[8,45,46,47,48], Joshua Pepper[49], Mark E. Rose[28], Jeffrey C. Smith[28,33], Keivan G. Stassun[50,51] & Liang Yu[3,52]

Astronomers have discovered thousands of planets outside the Solar System[1], most of which orbit stars that will eventually evolve into red giants and then into white dwarfs. During the red giant phase, any close-orbiting planets will be engulfed by the star[2], but more distant planets can survive this phase and remain in orbit around the white dwarf[3,4]. Some white dwarfs show evidence for rocky material floating in their atmospheres[5], in warm debris disks[6–9] or orbiting very closely[10–12], which has been interpreted as the debris of rocky planets that were scattered inwards and tidally disrupted[13]. Recently, the discovery of a gaseous debris disk with a composition similar to that of ice giant planets[14] demonstrated that massive planets might also find their way into tight orbits around white dwarfs, but it is unclear whether these planets can survive the journey. So far, no intact planets have been detected in close orbits around white dwarfs. Here we report the observation of a giant planet candidate transiting the white dwarf WD 1856+534 (TIC 267574918) every 1.4 days. We observed and modelled the periodic dimming of the white dwarf caused by the planet candidate passing in front of the star in its orbit. The planet candidate is roughly the same size as Jupiter and is no more than 14 times as massive (with 95 per cent confidence). Other cases of white dwarfs with close brown dwarf or stellar companions are explained as the consequence of common-envelope evolution, wherein the original orbit is enveloped during the red giant phase and shrinks owing to friction. In this case, however, the long orbital period (compared with other white dwarfs with close brown dwarf or stellar companions) and low mass of the planet candidate make common-envelope evolution less likely. Instead, our findings for the WD 1856+534 system indicate that giant planets can be scattered into tight orbits without being tidally disrupted, motivating the search for smaller transiting planets around white dwarfs.

WD 1856+534 (hereafter WD 1856 for brevity) is located 25 parsecs (pc) away in a visual triple star system. It has an effective temperature of 4,710 ± 60 K and became a white dwarf 5.9 ± 0.5 billion years ago (Gyr), based on theoretical models for how white dwarfs cool over time. The total system age, including the star's main-sequence lifetime, must be older. Tables 1 and 2 give the parameters of the system and the star, respectively. WD 1856 is one of thousands of white dwarfs that were targeted for observations with NASA's Transiting Exoplanet Survey Satellite (TESS), in order to search for any periodic dimming events caused by planetary transits. A statistically significant (12.1σ) transit-like event was detected by the TESS Science Processing Operations Center (SPOC) pipeline based on 28 days of data acquired between 18 July and 14 August 2019 (unless specified, all dates are given in UTC). The signal was rejected by an automated classification system designed to

A list of affiliations appears at the end of the paper.

Table 1 | Basic parameters of WD 1856+534

Other designations		
TIC 267574918		
TOI 1690		
LP 141–14		
2MASS J18573936+5330332		
Gaia DR2 2146576589564898688		
Astrometric parameters	**Value**	**Source**
Right ascension	18 h 57 min 39.34 s	Gaia
Declination	+53° 30′ 33.3″	Gaia
Right ascension proper motion	240.759 ± 0.148 mas yr^{-1}	Gaia
Declination proper motion	-52.514 ± 0.143 mas yr^{-1}	Gaia
Parallax	40.3983 ± 0.0705 mas	Gaia
Distance to star	24.754 ± 0.044 pc	Gaia
Literature and new photometric measurements		
g	17.6038 ± 0.0046	Pan-STARRS
r	16.9085 ± 0.0025	Pan-STARRS
i	16.6248 ± 0.0038	Pan-STARRS
z	16.5182 ± 0.0032	Pan-STARRS
y	16.4685 ± 0.0064	Pan-STARRS
G	16.9580 ± 0.0010	Gaia
B_P	17.5032 ± 0.0059	Gaia
R_P	16.2780 ± 0.0033	Gaia
J	15.677 ± 0.055	2MASS
H	15.429 ± 0.094	2MASS
K	15.548 ± 0.186	2MASS
W1	15.011 ± 0.027	ALLWISE
W2	15.156 ± 0.048	ALLWISE
W3	>13.404 (2σ)	ALLWISE
W4	>9.639 (2σ)	ALLWISE
Spitzer InfraRed Array Camera (IRAC) 4.5 μm	15.042 ± 0.066	This work

Reported uncertainties represent 68% confidence intervals (1σ) unless stated otherwise.
mas, milliarcsecond.

Table 2 | Derived stellar properties of WD 1856+534

Parameter	Value
Mass, M_*	$0.518 \pm 0.055 M_{\odot}$
Radius, R_*	$0.0131 \pm 0.00054 R_{\odot}$, $1.429 \pm 0.059 R_{\oplus}$
Surface gravity, $\log(g_{\mathrm{cgs}})$	7.915 ± 0.030
Effective temperature, T_{eff}	$4{,}710 \pm 60$ K
Cooling age, t_{cool}	5.85 ± 0.5 Gyr
Calcium abundance, log[Ca/(H + He)]	<−11.1
Iron abundance, log[Fe/(H + He)]	<−8.8
Magnesium abundance, log[Mg/(H + He)]	<−7.9
Sodium abundance, log[Na/(H + He)]	<−10.3
Sulfur abundance, log[S/(H + He)]	<−3.3

Reported uncertainties represent 68% confidence intervals (1σ) unless stated otherwise.
Values are from this work.
$M_{\odot}$, mass of the Sun; $R_{\odot}$, radius of the Sun; $R_{\oplus}$, radius of Earth.

identify planets around main-sequence stars. We noticed the signal in a visual inspection of all possible transit-like events detected around white dwarfs. As usual, caution is required when interpreting TESS data because of its coarse angular resolution; in this case, the white dwarf was blended together with several much brighter stars in the TESS images. However, the duration of the signal, approximately 8 min, is much shorter than the usual duration of ≳30 min for the transit of a main-sequence star, strongly suggesting that the transit signal originates from the white dwarf and not from the other stars.

To better characterize the transit signal, we obtained data with higher angular resolution. On 10 and 17 October 2019, we observed transits with three small privately operated telescopes, revealing that the white dwarf dims by up to 56% for 8 min. On 22 October 2019, we observed a transit with two larger telescopes, the Telescopio Carlos Sánchez and Gran Telescopio Canarias (Fig. 1). Together, these data show that a Jupiter-sized object transits the white dwarf in a grazing configuration (that is, the companion only occults part of the much smaller star).

Jupiter-sized objects can have a wide range of masses, ranging from giant planets (with masses as low as approximately $0.1M_J$; M_J, mass of Jupiter) to low-mass stars (around $100M_J$). Determining the mass is usually achieved through precise Doppler monitoring of the primary star. However, the spectrum of WD 1856 is classified[15] as type DC, a featureless continuum with no strong optical absorption or emission features. Optical and near-infrared spectra from MMT and the Lick Shane, Gemini North and Hobby–Eberly telescopes confirmed this classification (Fig. 2). The lack of strong spectroscopic absorption features precludes precise Doppler observations.

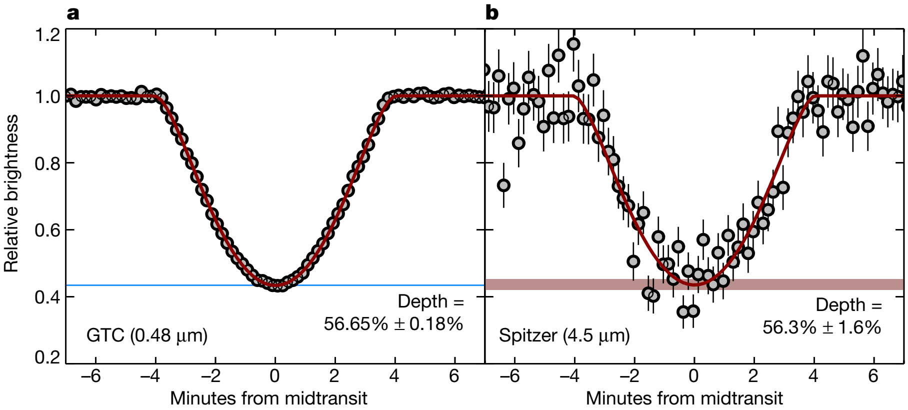


Fig. 1 | Transit observations of WD 1856. a, Optical transit observations with the Gran Telescopio Canarias (GTC). **b**, Infrared transit observations with the Spitzer Space Telescope. The red curves are the best-fitting models. The horizontal coloured shaded regions (light blue for GTC, light red for Spitzer) show the 68% confidence interval for the maximum loss of light. Any thermal emission from the transiting body would have led to a smaller loss of light at infrared wavelengths. The lack of any observed difference implies that the transiting body has a mass smaller than $13.8M_J$ (with 95% confidence). Each Spitzer point is an average of five exposures (each with a two-second exposure time), and the error bars show the 1σ error on the mean. The uncertainties on the GTC points are smaller than the size of the symbols.

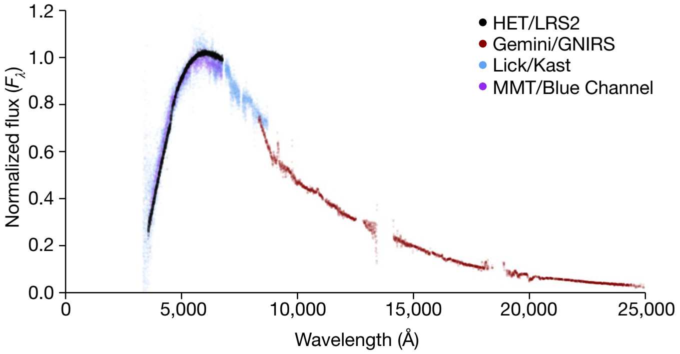


Fig. 2 | Spectroscopic observations of WD 1856. We show spectra from four observatories, using the Low Resolution Spectrograph 2 on the Hobby–Eberly Telescope (HET/LRS2), the Gemini Near InfraRed Spectrograph on the Gemini North Telescope (Gemini/GNIRS), the Kast Double Spectrograph on the Shane Telescope at Lick Observatory (Lick/Kast) and the Blue Channel spectrograph at MMT Observatory. The data have been scaled to remove offsets in their absolute flux calibrations. The optical spectra show a pure continuum, confirming the DC spectral classification, whereas the near-infrared spectrum from Gemini North shows only spurious features, owing to imperfect correction of the telluric absorption and sky emission from Earth's atmosphere. The units shown on the y axis are normalized flux per wavelength, λ.

Instead, we constrained the mass of the transiting body on the basis of the lack of any detectable thermal emission. We observed a transit on 16 December 2019 with NASA's Spitzer Space Telescope operating at wavelengths between 4 and 5 µm. At these infrared wavelengths, the thermal emission from a low-mass star or brown dwarf would make a larger fractional contribution to the total light than at the optical wavelengths of our other observations. This, in turn, would cause the fractional loss of light during transits to be smaller at infrared wavelengths than at optical wavelengths (absent slight differences in the stellar limb-darkening profile between the two bands). Figure 1 compares the infrared and optical light curves. There is no discernible difference in the fractional loss of light; any thermal flux from the transiting body can be no more than 6.1% of the flux from the white dwarf (with 95% confidence).

Such a faint object can only be a planet or a very-low-mass brown dwarf, based on theoretical models of brown dwarf evolution[16] and atmospheres[17]. Figure 3 shows the resulting constraints on the mass of the transiting companion as a function of the system age. A mass exceeding 13.8M_J is ruled out regardless of age (95% confidence), and the constraints for younger systems are even stronger. The system's motion through space suggests it is a member of the Galaxy's thin disk, implying an age less than about 10 Gyr and a mass less than 11.7M_J (95% confidence). Therefore, the transiting body almost certainly has a mass in the planetary regime[18].

Most or all of the usual circumstances that sometimes result in 'false positive' transiting exoplanet detections can be ruled out, given the data at hand. The ground-based transit observations confirm that the TESS signal is not an instrumental artefact or contamination from a different source. The transit duration is too long for the companion to be another white dwarf with an orbit of either 1.4 or 2.8 days. There is no evidence for unresolved blended sources in archival images or in the astrometric data from the Gaia mission of the European Space Agency (ESA). Even if there was a faint undetected companion, the transits are deep enough (>50%) that they must originate from WD 1856. Furthermore, the >50% transit depth implies that the signal also cannot be primary and secondary eclipses of an equal-temperature white dwarf/white dwarf binary. We conclude that WD 1856 is orbited by either a giant planet or a very-low-mass brown dwarf, which we designate WD 1856 b (for properties, see Table 3).

To avoid destruction when the progenitor of WD 1856 evolved into a red giant, WD 1856 b must have been further than about 1 AU from its host star, raising the question of how it arrived in the close orbit we observe today. Most short-period white dwarf binaries, including the small number of known white dwarf/brown dwarf pairs[19–22], are believed to have formed via common-envelope evolution[23]. In this theory, an expanding giant star grows large enough to engulf a lower-mass binary companion. Friction from the gaseous envelope of the giant star causes the companion to rapidly spiral inward towards the giant's dense core, depositing its orbital energy into the envelope. If the companion and core have enough gravitational potential energy, the envelope can be ejected, halting the orbital evolution of the companion and resulting in a binary system with an orbital period ranging from hours to days. If there is not enough gravitational potential energy to unbind the envelope, then the companion continues spiralling inwards towards the giant star's core until they merge.

It is difficult to explain the current orbit of WD 1856 b with standard common-envelope theory. Compared to a list[16] of known close white

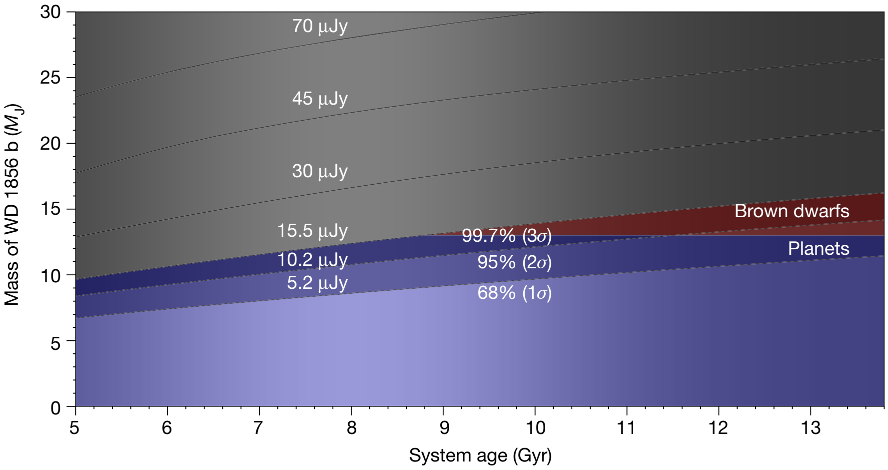


Fig. 3 | Allowed mass range for WD 1856 b as a function of the system age. Giant planets and brown dwarfs cool and contract as they age, and so higher masses are allowed by our Spitzer observations for older systems. The masses and ages comprising the grey region at the top of the plot (high masses) are excluded by the lack of any detectable thermal emission with Spitzer. The blue and red regions are the allowed ranges for planet and brown dwarf solutions, respectively, and are separated by the traditional 13M_J deuterium-burning limit[30]. The 1σ (68% confidence), 2σ (95%) and 3σ (99.7%) regions have darker shades, representing increasingly unlikely solutions. Several additional contours of constant brightness in the Spitzer 4.5-µm band are shown and labelled. To convey that the system's most probable age is ≲10 Gyr, the background has been shaded darker for much older ages.

Table 3 | Properties of planet candidate WD 1856+534 b

Parameter	Value	Value (eccentric fit)
Orbital period[a], P	1.4079405 ± 0.0000011 days	1.4079405 ± 0.0000011 days
Time of transit, t_t	2,458,779.3750828 ± 0.0000034 BJD_{TDB}	2,458,779.37508 ± 0.00012 BJD_{TDB}
Radius ratio, R_p/R_*	7.28 ± 0.65	$10.8^{+3.9}_{-2.6}$
Scaled semimajor axis, a/R_*	336 ± 14	325 ± 18
Semimajor axis, a	0.0204 ± 0.0012 AU	0.0198 ± 0.0014 AU
Orbital inclination, i	88.778 ± 0.059 degrees	$87.4^{+1.0}_{-1.7}$ degrees
Orbital eccentricity, e	0	<0.68 (2σ)
Transit duration, t_{14}	7.998 ± 0.023 min	7.945 ± 0.037 min
Planet radius, R_p	$(10.4 \pm 1.0)R_\oplus$	$15.4^{+5.5}_{-3.7}R_\oplus$
Transit impact parameter, b	7.16 ± 0.65	$10.7^{+3.9}_{-2.6}$
Incident flux, S	$(0.181 \pm 0.018)S_\oplus$	$0.212^{+0.041}_{-0.029}S_\oplus$
Equilibrium temperature[b], T_{eq}	163^{+14}_{-18} K	164^{+14}_{-18} K
Spitzer dilution parameter, d	0.004 ± 0.029	0.004 ± 0.028
Apparent IRAC 4.5 μm magnitude	>18.1 (2σ)	>18.1 (2σ)
Absolute IRAC 4.5 μm magnitude	>16.1 (2σ)	>16.2 (2σ)

[a] The reported orbital period is the value measured by observers in our Solar System's barycentric frame (that is, slightly Doppler shifted from the orbital period in the rest frame of the WD 1856 system).

[b] Equilibrium temperature T_{eq} calculated assuming an albedo a uniformly distributed between 0 and 0.7 and perfect heat redistribution: $T_{eq} = T_{eff}(1-a)^{1/4}\sqrt{\frac{R_*}{2a}}$.

The reported uncertainties represent 68% confidence intervals (1σ) unless stated otherwise. Values are from this work.

BJD_{TDB}, barycentric Julian date in barycentric dynamical time; $R_\oplus$, radius of Earth; $S_\oplus$, flux of Earth.

dwarf/brown dwarf binaries that were thought to have formed via common-envelope evolution, WD 1856 b has by far both the lowest mass and also the longest orbital period of any similar system. This implies that the gravitational potential energy released during the common-envelope phase is very small, which in turn makes it difficult to successfully eject the envelope of the WD progenitor. The amount of gravitational potential energy to be released is

$$\Delta\phi \approx -\frac{GM_{wd}M_{com}}{a} = \left(\frac{2\pi G}{P}\right)^{2/3}\frac{M_{wd}M_{com}}{(M_{wd}+M_{com})^{1/3}} \propto M_{com}(M_{wd}/P)^{2/3} \quad (1)$$

where M_{wd}, M_{com}, a and P are the white dwarf mass, the companion mass, the orbital separation and the orbital period, respectively, after the common-envelope phase. The brown dwarfs in the compiled systems[16] tend to have masses of at least $(50–60)M_J$ and orbital periods in the range of approximately 1–4 h. The low mass of WD 1856 b ($\lesssim 14M_J$) and long orbital period (about 34 h) could therefore have released only approximately 15 times less gravitational potential energy than the other systems listed[16]. More formally, we calculated that throughout most of the progenitor's giant phases, the gravitational potential energy release of WD 1856 b was insufficient to eject the envelope of the progenitor giant star and avoid merging with its core (see Methods). Some studies have suggested that the envelope's own internal energy could contribute to its ejection[24], but even this extra energy source appears insufficient for WD 1856 b to have ejected the envelope. WD 1856 b can probably only have formed by this mechanism if the common-envelope phase began after much of the envelope's mass had already been lost. Given the difficulty in forming WD 1856 b via common-envelope evolution and the degree to which it stands out from the population of known post-common-envelope binaries, we conclude that the system's current configuration most probably formed via some other mechanism.

Instead, a more probable formation history is that WD 1856 b was a planet that underwent dynamical instability. It is well established that when stars evolve into white dwarfs, their previously stable planetary systems can undergo violent dynamical interactions[13,25] that excite high orbital eccentricities. We have confirmed with our own simulations that WD 1856 b-like objects in multi-planet systems can be thrown onto orbits with very close periastron distances. If WD 1856 b were on such an orbit, the orbital energy would have rapidly dissipated, owing to tides raised on the planet by the white dwarf[26,27]. The final state of minimum energy would be a circular, short-period orbit. The advanced age of WD 1856 (around 5.85 Gyr) gives plenty of time for these relatively slow (of the order of Gyr) dynamical processes to take place. In this case, it is no coincidence that WD 1856 is one of the oldest white dwarfs observed by TESS.

Future observations should be able to confirm the planetary nature of WD 1856 b or—less likely—show that it is a low-mass brown dwarf. The amplitude of features in a planet's transmission spectrum depend inversely on the strength of its surface gravity. If WD 1856 b has a mass close to that of Jupiter, its spectral features could have amplitudes of about 1%. However, weak spectral features do not necessarily imply a large mass for WD 1856 b, because spectral features can be muted by high-altitude clouds or hazes[28]. Another path to measuring the mass of WD 1856 b would be to replicate our Spitzer observations with the upcoming James Webb Space Telescope (JWST). With its much larger collecting area, a single JWST transit observation should either detect thermal emission from WD 1856 b or place a strong enough constraint on its mass to confirm its planetary nature.

WD 1856 b will be a focus of future observational and theoretical studies. If the object's mass is low enough for it to cool to its equilibrium temperature (about 165 K), transmission spectroscopy observations could probe chemical species such as methane and ammonia in the atmosphere of one of the coldest known transiting planets[1]. If, instead, WD 1856 b has a higher mass and has retained some of its primordial heat, the low luminosity of the white dwarf means infrared observations with JWST could reveal the thermal emission spectrum of WD 1856 b with unusual detail. Regardless of its exact mass, WD 1856 b demonstrates that low-mass objects can migrate into close orbits around white dwarfs while avoiding total tidal disruption. Unlike common-envelope evolution—which predicts that low-mass objects will merge with the core of their host star—there is no reason why the dynamical mechanisms we invoke to explain the formation of WD 1856 b could not also be applied to even smaller planets, similar in size to Earth[29].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2713-y.

1. Akeson, R. L. et al. The NASA Exoplanet Archive: data and tools for exoplanet research. *Publ. Astron. Soc. Pacif.* **125**, 989–999 (2013).
2. Villaver, E. & Livio, M. The orbital evolution of gas giant planets around giant stars. *Astrophys. J. Lett.* **705**, 81–85 (2009).
3. Luhman, K. L., Burgasser, A. J. & Bochanski, J. J. Discovery of a candidate for the coolest known brown dwarf. *Astrophys. J. Lett.* **730**, 9 (2011).
4. Marsh, T. R. et al. The planets around NN Serpentis: still there. *Mon. Not. R. Astron. Soc.* **437**, 475–488 (2014).
5. Jura, M. A tidally disrupted asteroid around the white dwarf G29–38. *Astrophys. J. Lett.* **584**, 91–94 (2003).
6. Kilic, M., von Hippel, T., Leggett, S. K. & Winget, D. E. Excess infrared radiation from the massive DAZ white dwarf GD 362: a debris disk? *Astrophys. J. Lett.* **632**, 115–118 (2005).
7. Becklin, E. E. et al. A dusty disk around GD 362, a white dwarf with a uniquely high photospheric metal abundance. *Astrophys. J. Lett.* **632**, 119–122 (2005).
8. Gänsicke, B. T., Marsh, T. R., Southworth, J. & Rebassa-Mansergas, A. A gaseous metal disk around a white dwarf. *Science* **314**, 1908 (2006).

9. Wilson, T. G., Farihi, J., Gänsicke, B. T. & Swan, A. The unbiased frequency of planetary signatures around single and binary white dwarfs using Spitzer and Hubble. *Mon. Not. R. Astron. Soc.* **487**, 133–146 (2019).
10. Vanderburg, A. et al. A disintegrating minor planet transiting a white dwarf. *Nature* **526**, 546–549 (2015).
11. Manser, C. J. et al. A planetesimal orbiting within the debris disc around a white dwarf star. *Science* **364**, 66–69 (2019).
12. Vanderbosch, Z. et al. A white dwarf with transiting circumstellar material far outside the Roche limit. *Astrophys. J.* **897**, 171 (2020).
13. Debes, J. H. & Sigurdsson, S. Are there unstable planetary systems around white dwarfs? *Astrophys. J.* **572**, 556–565 (2002).
14. Gänsicke, B. T. et al. Accretion of a giant planet onto a white dwarf star. *Nature* **576**, 61–64 (2019).
15. McCook, G. P. & Sion, E. M. A catalog of spectroscopically identified white dwarfs. *Astrophys. J. Suppl. Ser.* **121**, 1–130 (1999).
16. Nelson, L., Schwab, J., Ristic, M. & Rappaport, S. Minimum orbital period of precataclysmic variables. *Astrophys. J.* **866**, 88 (2018).
17. Marley, M., Saumon, D., Morley, C. & Fortney, J. *Sonora 2018: Cloud-free, Solar Composition, Solar C/O Substellar Atmosphere Models and Spectra* (2018); https://doi.org/10.5281/zenodo.1309035
18. Spiegel, D. S., Burrows, A. & Milsom, J. A. The deuterium-burning mass limit for brown dwarfs and giant planets. *Astrophys. J.* **727**, 57 (2011).
19. Casewell, S. L. et al. WD0837+185: the formation and evolution of an extreme mass-ratio white-dwarf–brown-dwarf binary in Praesepe. *Astrophys. J. Lett.* **759**, 34 (2012).
20. Littlefair, S. P. et al. The substellar companion in the eclipsing white dwarf binary SDSS J141126.20+200911.1. *Mon. Not. R. Astron. Soc.* **445**, 2106–2115 (2014).
21. Rappaport, S. et al. WD 1202-024: the shortest-period pre-cataclysmic variable. *Mon. Not. R. Astron. Soc.* **471**, 948–961 (2017).
22. Parsons, S. G. et al. Two white dwarfs in ultrashort binaries with detached, eclipsing, likely sub-stellar companions detected by K2. *Mon. Not. R. Astron. Soc.* **471**, 976–986 (2017).
23. Paczynski, B. Common-envelope binaries. In *International Astronomical Union Symp. No. 73: Structure and Evolution of Close Binary Systems* (eds Eggleton, P., Mitton, S. & Whelan, J.) 75–80 (Reidel, 1976).
24. Xu, X.-J. & Li, X.-D. On the binding energy parameter λ of common-envelope evolution. *Astrophys. J.* **716**, 114–121 (2010).
25. Veras, D. & Gänsicke, B. T. Detectable close-in planets around white dwarfs through late unpacking. *Mon. Not. R. Astron. Soc.* **447**, 1049–1058 (2015).
26. Goldreich, P. & Soter, S. *Q* in the Solar System. *Icarus* **5**, 375–389 (1966).
27. Veras, D. & Fuller, J. Tidal circularization of gaseous planets orbiting white dwarfs. *Mon. Not. R. Astron. Soc.* **489**, 2941–2953 (2019).
28. Kreidberg, L. et al. Clouds in the atmosphere of the super-Earth exoplanet GJ1214b. *Nature* **505**, 69–72 (2014).
29. Agol, E. Transit surveys for Earths in the habitable zones of white dwarfs. *Astrophys. J. Lett.* **731**, 31 (2011).
30. Boss, A. P. et al. Working group on extrasolar planets. *Proc. International Astronomical Union A* **26A**, 183–186 (2005).

[1]Department of Astronomy, University of Wisconsin-Madison, Madison, WI, USA. [2]Department of Astronomy, The University of Texas at Austin, Austin, TX, USA. [3]Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]NSF's NOIRLab/Gemini Observatory, Hilo, HI, USA. [5]Department of Physics and Astronomy, University of Kansas, Lawrence, KS, USA. [6]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA. [7]Hereford Arizona Observatory, Hereford, AZ, USA. [8]Instituto de Astrofísica de Canarias (IAC), Tenerife, Spain. [9]Departamento Astrofísica, Universidad de La Laguna (ULL), Tenerife, Spain. [10]Los Alamos National Laboratory, Los Alamos, NM, USA. [11]Raemor Vista Observatory, Sierra Vista, AZ, USA. [12]Laboratory for Space Research, The University of Hong Kong, Hong Kong, China. [13]Center for Astrophysics and Space Sciences, University of California, San Diego, San Diego, CA, USA. [14]Center for Space and Habitability, University of Bern, Bern, Switzerland. [15]Max Planck Institute for Astronomy, Heidelberg, Germany. [16]Center for Astrophysics | Harvard & Smithsonian, Cambridge, MA, USA. [17]NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. [18]Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [19]Steward Observatory, University of Arizona, Tucson, AZ, USA. [20]Department of Physics and Astronomy, Dartmouth College, Hanover, NH, USA. [21]Department of Physics and Astronomy, University of California, Los Angeles, Los Angeles, CA, USA. [22]Department of Physics and Astronomy, Bishop's University, Sherbrooke, Quebec, Canada. [23]Hobby–Eberly Telescope, University of Texas, Austin, Austin, TX, USA. [24]DTU Space, National Space Institute, Technical University of Denmark, Kongens Lyngby, Denmark. [25]Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. [26]Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA. [27]Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA. [28]NASA Ames Research Center, Moffett Field, CA, USA. [29]Physics Department, University of Michigan, Ann Arbor, MI, USA. [30]Astronomy Department, University of Michigan, Ann Arbor, MI, USA. [31]Département de Physique, Université de Montréal, Montreal, Quebec, Canada. [32]Institut de Recherche sur les Exoplanètes (iREx), Université de Montréal, Montreal, Quebec, Canada. [33]SETI Institute, Mountain View, CA, USA. [34]Caltech/IPAC-NASA Exoplanet Science Institute, Pasadena, CA, USA. [35]Exoplanets and Stellar Astrophysics Laboratory (Code 667), NASA Goddard Space Flight Center, Greenbelt, MD, USA. [36]Noqsi Aerospace, Billerica, MA, USA. [37]Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, Los Angeles, CA, USA. [38]Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM, USA. [39]Department of Astronomy, University of California, Berkeley, Berkeley, CA, USA. [40]Department of Earth and Planetary Science, Graduate School of Science, The University of Tokyo, Tokyo, Japan. [41]Carl Sagan Institute, Cornell University, Ithaca, NY, USA. [42]Department of Astronomy and Space Sciences, Ithaca, NY, USA. [43]Department of Earth and Planetary Sciences, University of California, Riverside, Riverside, CA, USA. [44]Department of Physics and Astronomy, Moorpark College, Moorpark, CA, USA. [45]Astrobiology Center, Tokyo, Japan. [46]PRESTO, JST, Tokyo, Japan. [47]National Astronomical Observatory of Japan, Tokyo, Japan. [48]Komaba Institute for Science, The University of Tokyo, Tokyo, Japan. [49]Department of Physics, Lehigh University, Bethlehem, PA, USA. [50]Department of Physics and Astronomy, Vanderbilt University, Nashville, TN, USA. [51]Department of Physics, Fisk University, Nashville, TN, USA. [52]ExxonMobil Upstream Integrated Solutions, Spring, TX, USA.
✉e-mail: avanderburg@wisc.edu

Methods

TESS target selection and observations

We discovered the transits of WD 1856 b in data from NASA's TESS mission[31]. TESS is a satellite that observes a 96° × 24° region of sky with four 10-cm optical cameras. TESS observes the same region of sky continuously for approximately 28 days at a time; each 28-day observation is called a sector. Over the course of its two-year primary mission, TESS will observe 26 sectors, covering over 70% of the sky. TESS collects and downloads images of its entire field of view with 30-min exposure times, but TESS also observes 20,000 carefully chosen targets each month with shorter (two-minute) exposure times. Because transits of white dwarf stars typically have durations much shorter than the 30-min cadence of TESS's full-frame image downloads, we proposed two-minute-cadence observations of known and candidate white dwarf stars.

We proposed TESS observations of white dwarf stars in the Southern ecliptic hemisphere in late 2017, before the second data release (DR2) from ESA's Gaia mission enabled the discovery of hundreds of thousands of white dwarf candidates. We proposed two-minute cadence observations of white dwarfs in the Montreal White Dwarf Database (MWDD)[32] that are brighter than a magnitude of 17.5 in any of the V, I or TESS bands[33] and that are more than 20 arcsec from any brighter stars, which would contaminate the TESS photometric apertures. We also performed our own search (using the same V or I or TESS ≤ 17.5 magnitude limit) for, and proposed observations of, candidate white dwarfs by finding hot stars with high reduced proper motion (RPM)—a proxy for luminosity[34]. We used proper motions from the Hot Stuff for One Year catalogue[35], Gaia G-band magnitudes, and 2MASS J-band magnitudes to calculate each star's RPM. We defined cuts in colour/RPM space to select likely white dwarfs. A total of 615 unique white dwarf candidates from our program were observed during TESS's first year of operation.

In the second year of TESS observations of the northern ecliptic hemisphere, we identified targets from a catalogue of candidate white dwarfs[36] based on Gaia DR2. We proposed two-minute observations of all white dwarf candidates brighter than a Gaia G-band magnitude of 17 with a greater than 75% probability of being a true white dwarf, and removed white dwarfs less than 20 arcsec from any brighter stars, which would contaminate the TESS photometric apertures. Thanks to Gaia DR2, our northern target list was much more complete than our southern list. So far (as of sector 19), a total of 1,189 unique northern white dwarf candidates from our programme have been observed.

Once the TESS data on these targets were collected and downlinked from the spacecraft, they were processed by the Science Processing Operations Center (SPOC) pipeline[37,38] based at NASA Ames Research Center. The SPOC pipeline performed pixel-level calibrations, identified optimal photometric apertures, extracted light curves, corrected for systematic errors and diluting flux from nearby stars[39,40], and searched for periodic transit signals[41]. The periodic-transit search algorithm of the SPOC pipeline detected a convincing, 1.4-day-period, short-duration transit signal around WD 1856 (listed in the TESS Input Catalog as TIC 267574918). The transits were first detected in TESS's sector 14 observations, but the signal was rejected by an automatic classification algorithm designed to separate viable planet candidates from false positives (Guerrero, N. M. et al., submitted). We noticed WD 1856 in a visual inspection of all possible transit-like signals around white dwarfs identified by the SPOC pipeline (including those rejected by the automatic classifier), and initiated follow-up observations. Subsequently, WD 1856 was also observed in TESS sectors 15 and 19 (and will be observed again in sectors 22 and 26). The transits were re-detected in a combined analysis of the sector 14–15 data and in the sector 19 data. After being rejected by the automatic classifier in sectors 14 and 15, in the sector 19 observations the transit signal of WD 1856 b was promoted to the status of 'planet candidate' and given the designation TESS Object of Interest (TOI) 1690.01.

Although the TESS data confidently revealed the presence of 6–8-min-long, 1.4-day-period transits, and tests performed by the SPOC pipeline showed that the signal probably originated on WD 1856 (and not on some other nearby star), the TESS light-curve data were challenging to interpret. Compared to many other ground-based or space-based telescopes, TESS has relatively poor spatial resolution. The optics of TESS focus about 50% of a given star's light into one of its 20-arcsec pixels, and the wings of the point-spread function (PSF) extend several pixels farther. This poses challenges for observations of faint stars such as WD 1856, especially since it is only about 40 arcsec (2 pixels) away from a pair of physically associated M-dwarf stars (see below). The M-dwarfs are about 100 times brighter than WD 1856 in the TESS bandpass and contribute a substantial amount of flux into WD 1856's photometric aperture. In such situations, the dilution correction applied by the SPOC pipeline to the WD 1856 light curve is fairly uncertain given the difficulty in precisely measuring the wings of the TESS PSF. This uncertainty in the SPOC dilution correction translated to a substantial uncertainty in the true depth of the transits of WD 1856 b.

WD 1856 stands out among the stars in our TESS sample as one of the coolest—and therefore oldest—white dwarfs we observed. Among the 1,724 white dwarfs in our sample observed by TESS in sectors 1–19 with catalogue-reported effective temperatures[36], only eight white dwarfs are cooler than WD 1856.

Archival imaging and search for companions

We searched for both wide and close stellar companions to WD 1856 in archival survey data. WD 1856 was previously believed[15] to be part of a visual triple star system with a pair of M-dwarfs called G 229-20. G 229-20 consists of two nearly equal-brightness M-dwarf stars separated by about 2.3 arcsec (approximately 56 AU projected separation). The M-dwarf pair is located approximately 43 arcsec away from WD 1856 (approximately 1,000 AU projected separation). Data from Gaia DR2 show that G 229-20 A/B have nearly identical proper motions and parallaxes to WD 1856, confirming that the three stars are physically associated. From here on, we refer to the northern component of the binary as G 229-20 A since it is slightly brighter in resolved photometry from Gaia DR2.

We searched for additional co-moving companions in the Gaia archive. We queried all stars in Gaia DR2 within 600 arcsec of WD 1856 (approximately 15,000 AU projected separations) and looked for proper motions similar to that of WD 1856 and G 229-20 A/B. We found no stars with remotely similar space motions to that of the WD 1856 system.

We also checked to see if the Gaia observations showed any evidence for close, unresolved companions to either WD 1856 or G 229-20 A/B. Sometimes, close binary companions can introduce excess scatter into the Gaia astrometric observations[42,43]. This excess scatter is parameterized in a statistic called the renormalized unit weight error (RUWE)[44]. Solutions with low astrometric scatter have RUWE values close to 1, whereas stars with astrometric solutions that show anomalously high scatter (perhaps owing to astrometric motion from an unresolved binary companion) tend to have RUWE values greater than about 1.4. None of the members of the WD 1856 system show evidence for excess astrometric scatter that might reveal close companions; the RUWE values for WD 1856, G 229-20 A and G 229-20 B are 1.04, 1.01 and 0.94, respectively.

Finally, we searched for background stars at the present-day position of WD 1856 in archival imaging. WD 1856 was observed in the Palomar Observatory Sky Survey (POSS) on 27 July 1952 with a photographic plate with a blue-sensitive emulsion. Owing to its high proper motion, WD 1856 has moved over 16 arcsec since being imaged by POSS, making it possible to search for background stars at WD 1856's present-day position. There are no possible background contaminants at WD 1856's current position that are brighter than the POSS image's limiting magnitude (approximately 21st magnitude in blue)[45]. Extended Data Fig. 1 shows the POSS image of WD 1856 along with modern images from Pan-STARRS and TESS.

Ground-based transit follow-up

On the basis of the orbital period and time of transit inferred from the TESS observations of WD 1856, we planned ground-based transit observations to confirm the transit signal and measure its true depth. We observed transits of WD 1856 b on 10 October 2019 and 17 October 2019 with three small privately owned ground-based telescopes in Arizona: a 16-inch telescope at the Hereford Arizona Observatory (operated by B.G.), and a 16-inch telescope at Raemor Vista Observatory and a 32-inch telescope at Junk Bond Observatory (both operated by T.G.K.). We observed in white optical light without any colour filter; our effective bandpass was defined by the telescope systems' throughput and the CCDs' quantum efficiency. Weather conditions on both nights were clear and stable. The data were reduced following standard procedures for these telescopes[46]. All three telescopes confidently detected the transit signal with a consistent depth of around 60% on both nights. The data showed that the depths of odd- and even-numbered transits are indistinguishable and both greater than 50% of the total brightness, and so WD 1856 must not be a nearly equal-brightness eclipsing binary star with a true orbital period of 2.8 days, because the sum of the depths of the primary and secondary eclipses of a binary cannot exceed 100%.

After confirming the transits and determining the depth, we observed another transit of WD 1856 b with two larger telescopes to more precisely determine the transit shape and attempt to detect or rule out any colour dependence in the transit depth. We observed a transit of WD 1856 on 22 October 2019 with the MuSCAT2 instrument[47] on the 1.52-m Telescopio Carlos Sánchez and with the Optical System for Imaging and Low–Intermediate-Resolution Integrated Spectroscopy (OSIRIS) imager/spectrograph on the 10.4-m Gran Telescopio Canarias. MuSCAT2 provides simultaneous multi-colour images of a 7.4 × 7.4 arcmin field of view with fast readout times. We observed in four MuSCAT2 bands simultaneously: g, r, i and z_s. We reduced the observations with the standard MuSCAT2 pipeline and detected the transit with the same depth in each of the four MuSCAT2 bands. Our GTC observations used OSIRIS as an imager to obtain a precise g'-band light curve of WD 1856. We obtained 10-s exposures of WD 1856 and read out the detector in frame transfer mode, which allowed us to observe nearly continuously (one frame was read out while the next was exposing). We reduced the observations using standard Image Reduction and Analysis Facility (IRAF) scripts to calibrate the images and extract light curves for both WD 1856 and comparison stars. We experimented with different sized photometric apertures, and found that a six-pixel aperture minimized the scatter in the light curve. The resulting light curve was extremely precise (0.5% scatter per 10-s exposure) and revealed a smooth, symmetric 56% deep transit.

Our follow-up light curves are shown in Extended Data Fig. 2, compared to the TESS discovery light curve (corrected for the dilution from nearby stars).

Spectroscopy of WD 1856

A previous study assigned WD 1856 a spectral type classification of DC, indicating a continuum-dominated spectrum with very few weak absorption features[15]. We sought to confirm this classification and detect any weak absorption features by collecting our own optical spectroscopic observations. We observed WD 1856 on 5 October 2019 with the Blue Channel spectrograph[48] on the 6-m MMT telescope at Fred L. Whipple Observatory. We used the 500 line per mm grating and achieved 3.8-Å spectral resolution over a bandpass from 3,700–6,800 Å. A 10-min exposure yielded a signal-to-noise ratio of about 50 per pixel or 80 per resolution element. The resulting spectrum confirmed the DC spectral classification.

We continued searching for features in the spectrum of WD 1856 by extending our wavelength coverage beyond the red limits of our MMT Blue Channel observations. We obtained 60-min exposures of WD 1856 on 11 and 12 October 2019 with the Kast Double Spectrograph[49] on the 3-m Shane Telescope at Lick Observatory. On both nights, we configured the blue arm of the spectrograph to yield spectra with a resolving power $R = \lambda/\Delta\lambda = 1{,}300$ over the wavelength range 3,420–5,480 Å. We changed the configuration of the red arm between the two observations; on 11 October, we observed over a bandpass from 5,570 Å to 7,860 Å, and our 12 October observations pushed further red, from 6,400 Å to 8,800 Å (both with $R = 3{,}500$).

We observed WD 1856 on 30 October 2019 and 1 November 2019 with the Low Resolution Spectrograph 2 (LRS2)[50] on the 10-m Hobby–Eberly Telescope at McDonald Observatory. LRS2 is a combination of two integral field dual-channel spectrographs: one operating in the blue (3,700 Å to 7,000 Å) and one operating in the red (6,500 Å to 10,500 Å). We observed WD 1856 with the two blue channels of LRS2 with a spectral resolving power of $R = \lambda/\Delta\lambda = 1{,}910$ from 3,700–4700 Å and $R = 1{,}140$ from 4,700 Å to 7,000 Å. Each observation was 30 min in duration. The spectra were initially reduced with the automatic Hobby–Eberly Telescope pipeline, Panacea[51]. The pipeline performs basic CCD reduction tasks, wavelength calibration, fibre extraction, sky subtraction and flux calibration. We used the flux-calibrated, fibre-extracted spectra for the ultraviolet (3,700–4,700 Å) and orange channels (4,600–7,000 Å) to construct a single data cube correcting for differential atmospheric refraction and the small 0.3-arcsec offset between the two channels. We collapsed the data cube along the wavelength axis into an image of the LRS2 field of view, identified all fibres with at least 33% the flux of the brightest fibre, and summed the flux in those particular fibres at each wavelength in the data cube to extract a spectrum. The LRS2 spectra had the highest signal-to-noise ratio of all of our observations, but still showed no compelling evidence for any spectral features. In particular, the LRS2 spectra rule out any Hα absorption feature deeper than about 1%.

Finally, we observed WD 1856 on 21 November 2019 with the Gemini Near InfraRed Spectrograph (GNIRS)[52] on the 8.1-m Gemini North telescope (programme ID GN-2019B-DD-107) at Maunakea Observatory in Hawaii. The 32 lines per mm grating was used in the cross-dispersed mode, which provides continuous wavelength coverage from 1.0–2.5 μm. A slit width of 1.0 arcsec yielded a spectral resolving power of $R \approx 500$. Our total exposure time was 48 min, broken into 12 individual exposures (three sets of four exposures offset in an ABBA pattern). A telluric standard (HIP 95656) was observed immediately after the science observations. The observing conditions were excellent: sky was clear and seeing was about 0.35 arcsec in the H band around the target. Data reduction was performed using the XDGNIRS pipeline[53] v2.2.6. The correction for sky emission features and absorption due to Earth's atmosphere was imperfect and introduced some artefacts into the data, but we saw no evidence that any of the features in the data were spectral lines from the atmosphere of WD 1856. Our spectra of WD 1856 are shown in Fig. 2.

Spectroscopy of G 229-20 A/B

We also obtained ground-based optical spectra of G 229-20 A/B, the co-moving companion pair to WD 1856. We observed G 229-20 A/B with the Kast Double Spectrograph on the 3-m Shane Telescope at Lick Observatory. These observations were conducted on 11 October 2019, the same night as the first of our two Kast observations of WD 1856, and were taken with an identical instrument setting ($R = 1{,}300$ from 3,420 Å to 5,480 Å and $R = 3{,}500$ in the red from 5,570 Å to 7,860 Å). Seeing conditions were good enough to resolve the two stars, so we observed them simultaneously by rotating the spectrograph slit to the position angle of the binary and placing both stars on the slit. We extracted spectra of the two stars using standard IRAF routines. Although the stars were resolved, there was still some blending along the spatial axis.

We obtained medium-resolution spectra of G 229-20 A/B with two different echelle spectrographs. One spectrum came from the Fibre-fed Echelle Spectrograph (FIES)[54] on the Nordic Optical Telescope (NOT) on the island of La Palma, Spain on 18 February 2020. We used FIES in high-efficiency mode, in which the spectrogaph is fed with a 2.5-arcsec octagonal fibre to achieve a resolving power of $R = 25{,}000$. We reduced the spectra using the FIEStool pipeline[55]. We obtained the second spectrum with the Tillinghast Reflector Echelle Spectrograph (TRES)[56] on the 1.5-m telescope on Mount Hopkins, Arizona, USA on 24 February 2020. We used the standard instrumental setup with the spectrograph fed by a 2.3-arcsec fibre to achieve a spectral resolving power of $R = 44{,}000$. We reduced the spectra following standard practice for this instrument[57]. We cross-correlated the spectra with an archival observation of Barnard's Star and found that the absolute radial velocity of G 229-20 A/B is $17.9 \pm 0.1\ \mathrm{km\ s^{-1}}$ (by the IAU standard system)[58]. We also inspected the Hα line for G 229-20 A/B from the FIES spectrum. G 229-20 A/B have Hα in absorption, with an equivalent width of −0.32 Å (where equivalent width is defined to be positive for emission features).

We also used an archival spectrum of G 229-20 A published in a previous work[59]. The observation was made on 25 August 2006 with the MkIII spectrograph on the McGraw–Hill 1.3-m telescope at MDM Observatory and covered the wavelength range 6,200–8,700 Å. In that work, the authors assigned the star a spectral type of M3.5.

Spitzer observations

We observed a transit of WD 1856 b with the InfraRed Array Camera (IRAC) on NASA's Spitzer Space Telescope on 16 December 2019. We observed in IRAC channel 2, the reddest possible channel (sensitive to wavelengths of light between 4 and 5 μm) to best constrain the thermal flux from a faint, cool companion. We followed standard procedures for precise photometric observations with IRAC. We began with a 30-min-long 'burn-in' period during which we obtained dithered images of WD 1856 to allow both the spacecraft and detector to settle into equilibrium before the actual transit observations. We then observed WD 1856 for approximately two hours surrounding the predicted time of transit from our ground-based observations. These observations were conducted in 'peak-up' mode, in which WD 1856 was carefully placed on a well characterized pixel known to have minimal sensitivity variations. Images from a 32 × 32-pixel subarray were collected and saved every two seconds. Finally, after the transit observation was complete, we concluded our observations with 15 min of dithered imaging observations of WD 1856 for calibration purposes.

We analysed the Spitzer data with the Photometry for Orbits, Eclipses, and Transits (POET) pipeline[60]. POET extracts raw light curves from the images and optimizes a transit model while simultaneously modelling and removing spacecraft systematic errors. We investigated different sizes for the photometric aperture and found the best results with a small one-pixel radius (as expected for a star as faint as WD 1856). We optimized the transit and systematics model using a Markov chain Monte Carlo (MCMC) algorithm. The transit of WD 1856 was clearly detected in the Spitzer observations with nearly identical characteristics to the optical transit observations.

We also used the out-of-transit Spitzer observations to measure the combined flux of WD 1856 and WD 1856 b in IRAC channel 2. We measured the flux using standard aperture photometry as done in previous Spitzer observations of white dwarfs[61,62] using a two-pixel (1.2-arcsec) aperture (while applying a correction for any flux lost outside the aperture). We determined the total combined flux from WD 1856 and WD 1856 b in IRAC channel 2 to be 173 ± 10 μJy. We also searched for other faint red companions in the Spitzer observations. We coadded the individual Spitzer subarray observations to yield a deep 39 × 39-arcsec image of the region surrounding WD 1856 b. We detected one faint source (at RA = 18 h 57 min 39.9 s, dec. = +53° 30′ 48.9″), with a measured flux of 27 ± 5 μJy without an optical counterpart. Given its distance from WD 1856 (16 arcsec or 400 AU projected separation) and the M-dwarf companions (30 arcsec or 750 AU projected separation), we believe the source is more likely to be a background star or galaxy than a bound companion (since the probability of a chance alignment is high). Otherwise, we find no additional sources near WD 1856 with flux greater than 16 μJy (3σ confidence), which at the distance of the WD 1856 system corresponds to brown dwarfs with mass $m > 16M_\mathrm{J}$ (for ages up to 13.8 Gyr).

White dwarf stellar properties

We determined fundamental stellar parameters for WD 1856 (Table 2) using archival photometric observations and our high signal-to-noise optical spectra from the Hobby–Eberly Telescope. We followed the procedure of ref. [63] and fitted cool white dwarf spectral and evolutionary models[64] to broadband photometry from the Pan-STARRS and 2MASS surveys and the trigonometric parallax from Gaia DR2. We modelled the spectral energy distribution (SED)/spectra of WD 1856 with atmospheres with various compositions, ranging between $\mathrm{H/He} = 10^{-5}$ and $\mathrm{H/He} = 10^{2}$. We compared the predicted depth of the Hα absorption feature from the different models with the observed Hobby–Eberly Telescope spectrum (Extended Data Fig. 4); pure helium and most hydrogen/helium mixtures are consistent with our observed spectrum, but if WD 1856 b had a pure hydrogen atmosphere (or nearly so), we probably would have seen an Hα absorption feature in our Hobby–Eberly Telescope spectra. The models with at least some helium also were a better match to the observed SED; a pure hydrogen model over-predicts the near-infrared flux of WD 1856, whereas models with at least some helium better match the observations (see Extended Data Fig. 3).

We derived the white dwarf's fundamental stellar parameters from the results of our fits to the model atmospheres with varying ratios of hydrogen and helium. We found that a model with equal quantities of hydrogen and helium (50%/50% H/He) gave the best fit to the data. The resulting stellar parameters for some of the models we evaluated are given in Extended Data Table 1. The fits to pure hydrogen and 50%/50% H/He mixture yielded fairly consistent stellar parameters, whereas the pure helium atmosphere gave a considerably larger white dwarf and lower stellar mass. This discrepancy is due to the effects of He–He–He collision-induced absorption (CIA) in a pure helium atmosphere, which absorbs a substantial fraction of a white dwarf's infrared flux[65]. However, the efficiency of this opacity source is fairly uncertain, and it is plausible that its effects are overestimated in the pure He model.

We adopt the stellar parameters from the 50%/50% H/He model that best matched our observations and use them throughout the rest of the paper. However, the atmospheric composition of WD 1856 is not well constrained, and so we adopted conservative uncertainties on our stellar parameters. We inflated the formal uncertainties on the mass and radius from our model fits by adding a 10% and 3.3% uncertainty in quadrature, respectively. Our final adopted values for the mass and radius of the star are: $M_* = (0.518 \pm 0.055)M_\odot$ and $R_* = (0.0131 \pm 0.00054)R_\odot$ (Table 2).

We tested how much our results depend on the specific white dwarf models used by rederiving the stellar parameters of WD 1856 using alternate methods. We fitted[66] the SED of WD 1856 with a simple blackbody curve and found a best-fit temperature of $T_\mathrm{eff} = 4{,}720 \pm 50\ \mathrm{K}$, a bolometric flux $F_\mathrm{bol} = (3.93 \pm 0.23) \times 10^{-12}\ \mathrm{erg\ s^{-1}\ cm^{-2}}$ and a stellar radius of $R_* = (0.01298 \pm 0.00013)R_\odot$. Using an approximate fitting formula[67] designed to mimic the mass/radius relation from simple zero-temperature (black dwarf) models[68] and assuming a 2:1 oxygen/carbon ratio, we calculated a mass of $M_* = (0.54 \pm 0.01)M_\odot$. We also estimated the cooling age of WD 1856 using analytic relations[69] and found $t_\mathrm{cool} \approx 4$ Gyr, with uncertainties of roughly a factor of two[70]. All of these values are in good agreement with our adopted values, indicating that our results are fairly robust to different model assumptions.

Finally, we used the non-detection of spectroscopic features to place upper limits on the abundance of other elements in the atmosphere of WD 1856. With our Hobby–Eberly Telescope spectrum, we place strong limits on the presence of Ca, Fe, Mg and Na. When found in the atmospheres of white dwarfs, these elements are usually attributed to accretion from tidally disrupted rocky bodies such as asteroids or small planets. Because WD 1856 b is roughly the size of Jupiter, we also searched for elements more consistent with the composition of the atmosphere of a giant planet, such as those recently found[14] in the atmosphere of WD J0914+1914. It is harder to constrain the abundances of these elements because they show few spectral features at wavelengths covered by our spectroscopy. We can rule out sulfur abundances greater than log(S/H) = −3.3, but this limit is weaker than the measured sulfur abundance on WD J0914+1914. Future observations with higher spectral resolution and signal-to-noise will test whether WD 1856 shows evidence of accretion from its companion.

M-dwarf stellar properties

We determined the masses of G 229-20 A/B using broadband photometry and their Gaia DR2 trigonometric parallax measurements. In most photometric surveys (including 2MASS and Pan-STARRS), G 229-20 A and B are not well resolved and only have combined flux measurements. The two stars are, however, resolved in Gaia DR2 and have individually reported flux measurements. We converted the flux ratio of A/B from Gaia DR2 to a flux ratio in the 2MASS K-band using previously published spectrophotometric standards[71]. We then estimated the mass of each star using the previously published $M_{K_S} - M_*$relation[72], forcing the total K_S-band flux to match the unresolved measurement. This yielded masses of $(0.313 \pm 0.011)M_\odot$ and $(0.306 \pm 0.010)M_\odot$ for A and B, respectively. The unresolved 2MASS K_S measurement has a photometric-quality flag indicating a very poor profile fit (as expected for a close visual binary), so we also derived masses using the same method but without using the 2MASS measurement (and only the Gaia G-band magnitude), which yielded more conservative mass estimates of $(0.346 \pm 0.027)M_\odot$ and $(0.331 \pm 0.024)M_\odot$. We choose to adopt these more conservative estimates to avoid any possible systematic errors associated with the 2MASS data.

We checked these results for consistency by fitting[66] the SED of the two stars instead of empirical relations. Here, we fitted only the resolved Gaia G, B_P, and R_P magnitudes. We fixed the effective temperature of each M-dwarf to the values determined in the TICv8[73] ($T_{\mathrm{eff,A}} = 3{,}521$ K and $T_{\mathrm{eff,B}} = 3{,}513$ K) because those were already based on the resolved Gaia $B_P - R_P$ colours, and determined the bolometric flux of the two stars using the Gaia parallax. We determined the radii of the two stars to be $R_{*,A} = (0.35 \pm 0.02)R_\odot$, and $R_{*,B} = (0.34 \pm 0.02)R_\odot$. Converting from radii to masses using relations between the mass/radius of M-dwarfs and their absolute K-band magnitudes[71,72] yields $M_{*,A} = (0.335 \pm 0.024)M_\odot$, and $M_{*,B} = (0.322 \pm 0.023)M_\odot$. These results are in good agreement with our adopted masses.

Triple-system orbit analysis

We investigated the orbits of the three stellar components in the system comprising WD 1856 and G 229-20 A/B about the system's centre of mass. Gaia DR2 measured highly precise positions and proper motions for the three stars, so we used the Linear Orbits for the Impatient (LOFTI)[74] algorithm[75] to derive orbital constraints from these observations. Given input proper motions, positions, radial velocities (if available), and masses of the stellar components, LOFTI uses rejection sampling[76] to determine probability distributions for different orbital parameters.

We ran LOFTI to determine parameters for the orbits of WD 1856 and G 229-20 A/B about the centre of mass of the system. For the latter, we approximated G 229-20 A/B as a point mass. We used the masses determined in our earlier analysis, and ran LOFTI until the rejection-sampling algorithm had accepted 50,000 possible orbits. We found that the outer orbit is probably viewed close to face on (inclination, $i = 22^{+11}_{-11}$ degrees) and may be modestly eccentric ($0.30^{+0.19}_{-0.10}$). The semimajor axis is $a = 1{,}500^{+700}_{-240}$ AU, and the separation between WD 1856 and the centre of mass of G 229-20 A/B at closest approach is $a(1 - e) = 1{,}030^{+130}_{-55}$ AU.

We also ran LOFTI to determine parameters for the orbits of G 229-20 A and B about each other. Again, we ran the rejection sampler until we accumulated 50,000 samples in our posterior probability distribution. G 229-20 A and B orbit with a semimajor axis $a = 58^{+54}_{-16}$ AU and have a separation of $a(1 - e) = 39^{+27}_{-20}$ AU at their closest approach. The eccentricity of the orbit is not well constrained, with $e < 0.63$ (95% confidence) and the posterior probability distribution for the inclination peaks near 50 degrees ($i = 51^{+11}_{-17}$ degrees).

Transit analysis

We determined the best-fit values and uncertainties on the transit parameters and the flux of WD 1856 b at 4.5 μm with a simultaneous MCMC analysis of the GTC and Spitzer light curves. We first selected a small portion of both the Spitzer and GTC light curves near the observed transits; we used Spitzer data collected at times 2,458,834.27 ≤ BJD ≤ 2,458,834.30 and GTC data from 2,458,779.369 ≤ BJD ≤ 2,458,779.382 (after converting the GTC timestamps to $\mathrm{BJD_{TDB}}$)[77]. For convenience, we downsampled the two-second-cadence Spitzer light curve by a factor of five to match the 10-second cadence of the GTC light curve points. We divided the Spitzer and GTC data by the median out-of-transit flux measurement to set the out-of-transit flux level to 1. We estimated uncertainties on each point in the light curves by multiplying a value for the out-of-transit scatter (from the standard deviation of the normalized out-of-transit points) by the square root of each flux value.

We fitted the transits with exact analytic transit light-curve models[78] for stars with quadratic limb-darkening laws coupled to a code for solving Kepler's equation[79] (for fits with non-zero eccentricity). We oversampled the model light curves by a factor of six and integrated to account for the 10-second exposure time of both the GTC observations and our binned Spitzer observations. We fixed the limb-darkening parameters for the white dwarf to values calculated from model atmospheres. For our GTC g′-band observation we used coefficients specifically calculated for white dwarfs[80]. These coefficients, $u_1 = 0.05$ and $u_2 = 0.52$, closely match other independently calculated coefficients[81], $u_1 = 0.07$ and $u_2 = 0.46$. For our Spitzer observation we used coefficients from models of main-sequence stars with the same effective temperature[82], $u_1 = 0.0$ and $u_2 = 0.15$. We modelled the flux contribution of WD 1856 b (if any) to the Spitzer light curve by fitting for a dilution term, $d \equiv F_{\mathrm{WD1856b}}/F_{\mathrm{WD1856}}$. We calculated and re-normalized the Spitzer transit model $M_S(t)$ from the un-diluted transit model $M(t)$:

$$M_S(t) = \frac{M(t) + d}{1 + d}. \qquad (2)$$

At each MCMC link, we subtracted the transit models from the GTC and Spitzer light curves, fitted a quadratic polynomial to the residual light curves and added this polynomial curve to the transit model. This step marginalizes over any possible trends and normalization errors in the two light curves. We fitted for two additional photometric error terms (one for GTC and one for Spitzer) added in quadrature to our calculated uncertainties and imposed a Gaussian prior on the density of WD 1856 centred at 324,000 g cm^{-3} with a width of 54,000 g cm^{-3} based on our stellar parameters. Our knowledge of the stellar density allows us to calculate the average orbital speed of WD 1856 b via Kepler's third law[83] and to link the transit duration (a direct observable quantity) to the radius of the planet candidate. This information, along with a constraint on the transit impact parameter from the maximum depth of the transit, helps the MCMC converge to a well behaved solution.

The transit of WD 1856 is grazing, so even when imposing a prior on the white dwarf's stellar density, the radius of the transiting object is almost completely degenerate with the object's orbital speed at the time of transit. We therefore performed one fit assuming a circular orbit and another fit allowing for orbital eccentricity. When we assumed circular orbits, we fitted for 10 free parameters: orbital period, time of transit, cosine of the orbital inclination ($\cos i$), scaled semimajor axis (a/R_*), planet–star radius ratio (R_p/R_*), photometric jitter terms for both the Spitzer and GTC light curves, and the Spitzer dilution parameter d. Other than our prior on stellar density (which mostly affects a/R_*), we used uniform priors with bounds $(-\infty, \infty)$ on all parameters except for the jitter terms, a/R_*, R_p/R_*, which we restricted to $[0, \infty)$, and $\cos i$, which we restricted to $[0, 1]$. We did not force the dilution parameter to be positive to avoid a Lucy–Sweeney-like[84] bias. We explored parameter space with an affine invariant MCMC sampler[85] with 50 walkers evolved for 200,000 steps (discarding the first half for burn-in).

For our fits allowing eccentric orbits, we changed our parameterization to speed the MCMC convergence. Instead of exploring parameter space in $\cos i$, we defined a new parameter $\delta \equiv R_p/R_* - b$ (where $b = (a/R_*) \cos i$ is the transit impact parameter), to avoid a strong correlation between R_p/R_* and b. We also fitted for combinations of eccentricity e and argument of periastron ω ($\sqrt{e}\sin\omega$ and $\sqrt{e}\cos\omega$) for a similar reason. We imposed a physical cutoff for high eccentricity orbits; at each link, we calculated the instantaneous Roche lobe radius[86] of WD 1856 b at periastron R_L:

$$R_L \approx 0.46(1-e)a\left(\frac{M_p}{M_*}\right)^{1/3}, \tag{3}$$

assuming a planet mass $M_p = 15M_J$ (see below). We discarded any links where the planet's size exceeded this radius, which prevented the fit from diverging towards high eccentricities and large companion radii. Even with these modifications, the eccentric fit was much slower to converge; we evolved 50 walkers for 8,000,000 links, discarding the first 5,000,000 to remove the burn-in phase and save disk space. Correlations between selected parameters for both the circular and eccentric fits are shown in Extended Data Figs. 5, 6.

Both fits showed that WD 1856 b is a roughly Jupiter-sized object. If its orbit is circular, WD 1856 b has a radius $R_p = (10.4 \pm 1.0)R_\oplus$ ($R_\oplus$, radius of Earth); if eccentric orbits are allowed, the uncertainty on the radius is much larger: $R_p = 15.4^{+5.5}_{-3.7}R_\oplus$. Radii smaller than about $7R_\oplus$ are strongly ruled out in both cases, so the companion cannot be another white dwarf. Our fits also revealed that the transit depth at 4.5-μm wavelengths is nearly identical to the optical transit depth. We measure the Spitzer dilution parameter $d = 0.004 \pm 0.029$. Evidently, the flux of WD 1856 b is only a small fraction of the white dwarf itself at 4.5 μm. This places strong constraints on the temperature (and therefore mass) of WD 1856 b, as described below.

In principle, using inaccurate limb-darkening coefficients in our fits can adversely affect our measurement of the dilution coefficient and planet radius. We tested the robustness of our results to such errors by running additional MCMC fits where the limb-darkening coefficients were free parameters constrained by basic physical priors[87]. We ran three separate fits: one in which the Spitzer limb-darkening coefficients were restricted to probable values ($u_1 < 0.2$, $u_2 < 0.3$)[82] and the GTC coefficients were fixed to model values; one in which the Spitzer coefficients were free and the GTC coefficients were fixed to the model values; and one in which both the GTC and Spitzer limb-darkening coefficients were free. Our results are insensitive to the limb-darkening coefficients; our fit with the Spitzer coefficients restricted to ($u_1 < 0.2$, $u_2 < 0.3$) and GTC coefficients fixed to model values gave statistically identical results to our baseline fit. Even when both the Spitzer and GTC coefficients were allowed to freely vary, the dilution parameter and R_p/R_* shifted by only 0.2σ and 0.4σ, respectively.

Mass limit of the companion

We quantified the constraints placed by our Spitzer observations using brown dwarf/giant planet evolution and atmosphere models. From our measurement of $d = F_{\mathrm{WD1856b}}/F_{\mathrm{WD1856}}$ at 4.5 μm from our transit fits, and our measured total flux of WD 1856 and WD 1856 b at 4.5 μm (173 ± 10 μJy), we calculate the flux of WD 1856 b at 4.5 μm:

$$F_{\mathrm{WD1856\,b}} = dF_{\mathrm{WD1856}} = \frac{F_{\mathrm{total}}}{1 + 1/d} = 0.7 \pm 4.9\,\mu\mathrm{Jy} \tag{4}$$

When we exclude all unphysical solutions where $d < 0$, we calculate 68%, 95% and 99.7% upper limits on F_{WD1856b} at 4.5 μm that are 5.2, 10.2 and 15.5 μJy, respectively. We emphasize that this limit on the flux of WD 1856 b at 4.5 μm is model independent and does not rely on our white dwarf stellar parameters or SED fit.

We used the Sonora grid[17] of cloud-free solar metallicity brown dwarf/giant planet models to relate the thermal flux at 4.5 μm to atmospheric parameters such as effective temperature and surface gravity. We interpolated the predicted thermal flux in IRAC channel 2 from the Sonora atmosphere models onto two sets of evolutionary models: the underlying models used in the Sonora atmosphere calculations, and a more densely sampled grid of models[16] produced using the Modular Experiments in Stellar Evolution (MESA) code. We found that the two evolutionary grids gave nearly identical results, and adopted the MESA models given their denser sampling.

The MESA brown dwarf models predict the properties of objects with masses from $2.1M_J$ to $104M_J$ over 20 Gyr of evolution and are sampled at a total of 329,732 points in the mass–age plane. We compared the predicted 4.5 μm flux for each of these model points to determine the allowed brown dwarf masses given our constraints. We assume that WD 1856 b must be at least as old as the white dwarf's cooling age (roughly 5.85 Gyr) and cannot be older than the age of the universe (13.8 Gyr), so we ignore any model points outside this age range. We found that for the oldest (13.8 Gyr) possible brown dwarfs/giant planets, we constrain the mass to be less than $11.1M_J$ at 68% confidence (1σ), $13.8M_J$ at 95% confidence (2σ) and $16.1M_J$ at 99.7% confidence (3σ). The object's temperature must be below 250 K, 290 K or 320 K at 1σ, 2σ and 3σ confidence, respectively.

The tail of WD 1856 b's allowed mass distribution straddles the $13M_J$ deuterium-burning limit traditionally used to distinguish giant planets and brown dwarfs[30,88,89]. However, using the deuterium-burning limit to distinguish planets and brown dwarfs is imprecise. There is probably no specific mass above which deuterium burning takes place in brown dwarfs;[18] instead, the limit probably spans a range from about $11M_J$–$16M_J$ (depending on the object's composition and how the onset of deuterium burning is defined). It may also be more appropriate to divide planets and brown dwarfs by their formation histories[90,91]. Given the lack of a clear division between planets and brown dwarfs, we refer to WD 1856 b as a planet candidate until future observations can place stronger constraints on its mass.

These upper limits on the mass of WD 1856 b are model dependent, so we tested how they change when we use different model grids and assumptions. We repeated our calculation using the recently developed ATMO 2020 evolutionary and atmospheric models[92]. Because these models were only calculated to an age of 10 Gyr, we compared the 1σ, 2σ and 3σ upper mass limits with those for 10-Gyr objects with the Sonora and MESA models. We found good agreement in the mass upper limits between the two models (within about $2M_J$, with ATMO 2020 models yielding a lower 1σ mass limit and a higher 3σ mass limit, owing to stronger dependence of 4.5-μm flux on mass). Using the ATMO 2020 models, we also tested the effects of non-equilibrium chemistry, which can be important for cold brown dwarfs[93]. Even strong disequilibrium chemistry (with the vertical eddy diffusion coefficient $\log K_{zz} \approx 6.5$, where z is the direction away from the centre of the brown dwarf) had a minimal effect on our mass limits.

The effect of clouds on our mass limits is more difficult to quantify. In general, the presence of clouds slows the cooling of brown dwarfs and giant planets, and so objects with clouds should generally remain hotter and more luminous throughout their evolution[88]. However, when clouds are present, they can substantially change the object's spectrum and tend to decrease the flux in the 4.5-μm band[94]. Water clouds are expected to form in giant planets and brown dwarfs cooler than about 375 K (ref. [95]), so in the case of WD 1856 b, these two effects will probably compete. Future modelling should more fully reveal which effect dominates.

Age of the WD 1856 system

Giant planets and brown dwarfs cool as they age, and so our mass limits are stronger for younger systems. We therefore attempted to place additional constraints on the total system age in addition to the white dwarf cooling age (≳5.85 Gyr) and the age of the universe (<13.8 Gyr). One possible way to measure the age of a white dwarf is to add the white dwarf's cooling age to the estimated main-sequence lifetime of its progenitor star using a white dwarf initial–final mass relation. Unfortunately, two factors make it difficult to estimate the age of the progenitor. First, the white dwarf initial–final mass relations assume that the star evolved as an isolated single star and did not undergo mass transfer or a common-envelope phase. As we show below, although it is difficult, it is not impossible that WD 1856 b reached its current orbit by this mechanism. Second, a white dwarf progenitor's lifetime is a sensitive function of the white dwarf's final mass; a 50% increase in a white dwarf's mass from $0.5M_\odot$ to $0.75M_\odot$ corresponds to a 275% increase in the progenitor's mass, from $0.8M_\odot$ to $3M_\odot$, and a corresponding decrease in the star's main-sequence lifetime by a factor of approximately 20 (from about 10 Gyr to about 500 Myr). With a mass of $0.52M_\odot$, the white dwarf initial–final mass relation favours a long-lived progenitor with a mass less than that of the Sun and a total system age at least 15 Gyr, older than the age of the universe. Because our white dwarf model spectra struggle to describe our observations (see above), we suspect that systematic errors in our estimate of the mass of WD 1856 probably explain the system's apparently unphysical age. If the true mass were closer to $0.6M_\odot$ (only about 1.5σ away given our conservative uncertainties), this tension would disappear. We conclude that given these uncertainties, estimating the lifetime of WD 1856's progenitor cannot give a reliable system age.

We then shifted our attention to the binary M-dwarf pair G 229-20 A/B. Presumably these stars formed together with WD 1856's progenitor, and therefore should be the same age as WD 1856's planet candidate. It is notoriously difficult to determine the age of old (≳1 Gyr) field stars, and especially difficult for M-dwarfs, but there are some indicators that can broadly suggest an age for the system. We saw no evidence that the M dwarfs are particularly young; the two stars do not have Hα in emission, and light curves of the two stars from TESS, the All-Sky Automated Survey for Supernovae (ASAS-SN)[96,97] and the Super Wide Angle Search for Planets (SuperWASP) survey[98] show no evidence for a rotational variability. This is unsurprising since we assume G 229-20 A/B must have formed before WD 1856 became a white dwarf about 5.85 Gyr ago. However, we also saw no evidence that G 229-20 A/B are particularly old. Similar to most typical field age M dwarfs, the spectra of G 229-20 A/B show a band of prominent calcium hydride (CaH) and titanium oxide (TiO) absorption features[99] often characterized using the $\zeta_{\mathrm{TiO/CaH}}$ parameter[100,101]; if G 229-20 A/B were old sub-dwarfs, we would expect $\zeta_{\mathrm{TiO/CaH}} < 0.8$, but the value is 0.93, consistent with most Solar-metallicity M dwarfs. The Hα equivalent width (a proxy for magnetic activity and therefore age[102]) of G 229-20 A/B is lower than average, but still well within typical ranges for field M dwarfs[103] (see Extended Data Fig. 7).

We also investigated the system's galactic kinematics. Using the system's position, proper motion and parallax from Gaia DR2, along with our measured radial velocity (with an inflated uncertainty to account for the motion of the M dwarfs about the system barycentre), we calculated the system's three-dimensional space motion to be $(U, V, W) = (8.65 \pm 0.21, 40.4 \pm 1.8, -15.13 \pm 0.70)$ km s^{-1} with respect to the local standard of rest (LSR)[104]. We calculated the relative probabilities[105,106] that the WD 1856 system is a member of the galactic thin disk, thick disk or halo, and found that WD 1856 is most likely (93%) a member of the thin disk, with only about a 7% chance that it is part of the thick disk. Halo membership is strongly disfavoured (4,000:1 odds against). The mean age for stars in the thin disk[107] is about 7–8 Gyr (with large spread), and the oldest stars in the thin disk are probably around 8–10 Gyr in age[108,109]. Thick-disk stars are about 1.5–2 Gyr older on average than thin-disk stars, with a mean age[107,110] of approximately 9–10 Gyr.

All in all, these lines of evidence point to a system that is fairly old, but probably not much older than about 10 Gyr. If we assume the system is no older than 10 Gyr, the mass of WD 1856 b must be less than $9.4M_\mathrm{J}$, $11.9M_\mathrm{J}$ and $13.6M_\mathrm{J}$ at confidence levels of 1σ, 2σ and 3σ, respectively.

Common-envelope evolution

When WD 1856's progenitor star was in its main sequence, the companion WD 1856 b must have orbited farther from the progenitor than it does today, or it could not have survived the progenitor's red giant evolutionary phase. Here, we consider how WD 1856 b might have reached its current orbit close to WD 1856. One obvious possibility for placing a massive planetary object in a relatively close orbit with a white dwarf is common-envelope evolution[23,111,112]. Previous work[16] has investigated the likelihood that short-period, detached binaries containing a brown dwarf (or low-mass M dwarf) companion in orbit with a white dwarf (or hot subdwarf) could have been formed via a common-envelope phase of evolution. That work compiled a table of 25 binaries with orbital periods between 68 min and 4 h and showed that the measured masses of the companions—which typically fall in the range $50M_\mathrm{J}$–$100M_\mathrm{J}$—are not inconsistent with the predictions of common-envelope evolution. There are some detached systems that have orbital periods longer than 4 h with companion masses in this range, but none that we are aware of with periods as long as that of WD 1856 (1.4 days). Nonetheless, we will now examine whether it is possible for a $15M_\mathrm{J}$ object (at the upper end of our allowed mass distribution) to eject the envelope of a low-mass giant and end up in an orbit as long as 1.4 days.

There are a number of different ways to formulate the initial–final orbital separation, $a_\mathrm{f} - a_\mathrm{i}$, during a common-envelope phase, on the basis of conservation of energy. The fundamental idea is to determine the final orbital separation of the binary once the low-mass companion has ejected the common envelope of the progenitor, in terms of the initial orbital separation of the primordial binary and its constituent masses. More recent treatments of the energy formulation take into account the fraction of the internal energy used to eject the envelope, for example the recombination energy[24,113–115]. Conservation of energy relates a_f to a_i as follows:

$$\frac{GM_\mathrm{p}M_\mathrm{e}}{\lambda_\mathrm{CE}r_\mathrm{L}a_\mathrm{i}} = \alpha\left[\frac{GM_\mathrm{c}M_\mathrm{s}}{2a_\mathrm{f}} - \frac{GM_\mathrm{p}M_\mathrm{s}}{2a_\mathrm{i}}\right], \qquad (5)$$

where M_p and M_s are the masses of the primordial primary (the WD 1856 progenitor) and the primordial secondary star (in this case the massive planet candidate), respectively, M_c and M_e are the masses of the core and envelope of the primary star[111,116,117] and G is the gravitational constant. The parameter λ_CE^{-1} is a measure of the total gravitational binding energy of the envelope to itself and to the core of the primary star in units of $-GM_\mathrm{p}M_\mathrm{e}/R_\mathrm{p}$ and α is an energy-efficiency parameter for ejecting the common envelope. The factor $r_\mathrm{L} \equiv R_\mathrm{L}/a_\mathrm{i}$ is the dimensionless radius of the Roche lobe of the primary star when mass transfer commences. If the internal energy (for example, electron recombination) is taken into account, then either α or λ_CE may be considered to be larger than unity[24,115,118].

For the masses and separations relevant to the formation of the WD 1856 system, the second term in square brackets in equation (5) is negligible compared to the first term (see ref. [119] for a more detailed analysis). Upon dropping that term, we find:

$$\frac{a_{\mathrm{f}}}{a_{\mathrm{i}}} \approx \frac{\lambda_{\mathrm{CE}}\alpha r_{\mathrm{L}}}{2}\left(\frac{m_{\mathrm{c}}m_{\mathrm{s}}}{m_{\mathrm{e}}m_{\mathrm{p}}}\right), \tag{6}$$

where lowercase masses m are implicitly expressed in solar masses $M_{\odot}$. In turn, this can be expressed as the ratio of final-to-initial orbital periods:

$$\frac{P_{\mathrm{f}}}{P_{\mathrm{i}}} \approx \left(\frac{\lambda_{\mathrm{CE}}\alpha r_{\mathrm{L}}}{2}\right)^{3/2}\left(\frac{m_{\mathrm{c}}m_{\mathrm{s}}}{m_{\mathrm{e}}m_{\mathrm{p}}}\right)^{3/2}\left(\frac{m_{\mathrm{p}}+m_{\mathrm{s}}}{m_{\mathrm{c}}+m_{\mathrm{s}}}\right)^{1/2}. \tag{7}$$

The mass of the degenerate core of low-mass stars is closely related to the radius of the giant, and so it also follows that there is a relation between the orbital period and giant's core mass when mass transfer commences.

We illustrate the $R(M_{\mathrm{c}})$ relation in Extended Data Fig. 8. Here we show MIST[120] evolution tracks for solar metallicity stars in the radius–core mass plane. These are for seven different initial stellar masses covering a range $1.0M_{\odot}$–$2.8M_{\odot}$. On the first red giant branch there is a common locus of upper limits to the radius, whereas on the asymptotic giant branch (AGB) the same is true; the main difference is the thermal pulses, during which the radius varies substantially. The lime green curve superposed on the plot is an analytic expression that represents fairly well the locus of upper limits—which is where mass transfer to a companion star would first occur. The expression

$$R(m_{\mathrm{c}}) \approx \left(5.56\times10^{4}\frac{m_{\mathrm{c}}^{19/3}}{1+20m_{\mathrm{c}}^{3}+10m_{\mathrm{c}}^{6}}+4\right)R_{\odot}, \tag{8}$$

(for $0.7M_{\odot} \gtrsim M_{\mathrm{c}} \gtrsim 0.15M_{\odot}$) is modelled after equation (5) in ref. [121] and inferred from equation (12) in ref. [122], with some minor modifications.

The orbital period that corresponds to a primary with core mass m_{c} and which is just filling its Roche lobe with the secondary star is:

$$P_{\mathrm{i}} \approx 1.53\times10^{6} f(m_{\mathrm{c}})^{3/2}\frac{1}{r_{\mathrm{L}}^{3/2}}\frac{1}{\sqrt{m_{\mathrm{p}}+m_{\mathrm{s}}}}\ \mathrm{days}, \tag{9}$$

with

$$f(m_{\mathrm{c}}) \equiv \frac{m_{\mathrm{c}}^{19/3}}{(1+20m_{\mathrm{c}}^{3}+10m_{\mathrm{c}}^{6})}+f_{0}$$

where $f_0 = 7.2\times10^{-5}$. Here r_{L} has the same meaning as in equations (6) and (7).

We now combine equations (7) and (9) into a single equation for the post-common-envelope period, P_{pce}, and associate the system masses in equation (7) with those we observe in WD 1856: $m_{\mathrm{c}} \equiv m_{\mathrm{wd}}$, $m_{\mathrm{s}} \equiv m_{\mathrm{com}}$ and $m_{\mathrm{e}} \equiv m_{\mathrm{p}} - m_{\mathrm{c}}$, where the subscript 'com' represents the current companion to the white dwarf, which we believe is a gas-giant planet.

$$P_{\mathrm{pce}} \approx 5.4\times10^{5}(\lambda_{\mathrm{CE}}\alpha)^{3/2}\frac{f^{3/2}(m_{\mathrm{wd}})}{(m_{\mathrm{wd}}+m_{\mathrm{com}})^{1/2}}\left[\frac{m_{\mathrm{wd}}m_{\mathrm{com}}}{(m_{\mathrm{p}}-m_{\mathrm{wd}})m_{\mathrm{p}}}\right]^{3/2}\ \mathrm{days} \tag{10}$$

Note that the period of the post-common-envelope system is a function only of the masses of the companion, the white dwarf and its progenitor.

For the WD 1856 system we know that $P_{\mathrm{f}} = 1.4$ d, $M_{\mathrm{wd}} = 0.52M_{\odot}$ and we take $M_{\mathrm{com}} \approx 0.015M_{\odot}$ as an upper limit on the mass of the current companion object. Thus, we can use equation (10) to find the required value of $\alpha\lambda_{\mathrm{CE}}$ as a function of the primary mass (of the progenitor of the white dwarf):

$$\alpha\lambda_{\mathrm{CE}} \simeq 1.5\times10^{-4}P_{\mathrm{pce}}^{2/3}\frac{1}{f(m_{\mathrm{wd}})}\frac{(m_{\mathrm{wd}}+m_{\mathrm{com}})^{1/3}(m_{\mathrm{p}}-m_{\mathrm{wd}})m_{\mathrm{p}}}{m_{\mathrm{wd}}m_{\mathrm{com}}}. \tag{11}$$

Finally, in Extended Data Fig. 9 we plot equation (11) as a function of the mass of the primary progenitor star of the current white dwarf. From this figure we can see that for progenitor masses of $1M_{\odot}$, $2M_{\odot}$ and $3M_{\odot}$, values of the parameter $\alpha\lambda_{\mathrm{CE}} = 2.4$, 15 and 38 would be required to unbind the envelopes, respectively. According to a previous work[24] the calculated values of $\alpha\lambda_{\mathrm{CE}}$, including internal energies are $\lesssim 0.4$, $\lesssim 2$ and $\lesssim 5$, respectively (when the stellar radii are in the relevant range $100R_{\odot}$–$250R_{\odot}$, considerably less than the values required for WD 1856 b to eject the primary star's envelope. Without invoking internal energy, it appears even more improbable that a $15M_{\mathrm{J}}$ object could unbind the common envelope of the white dwarf's progenitor.

We explored whether WD 1856 b could have plausibly ejected a common envelope at any point in its progenitor's evolution by directly calculating the required $\alpha\lambda_{\mathrm{CE}}$ value from the MIST tracks. At each point in the MIST tracks at which the primary star was expanding to engulf new regions of its solar system, we calculated the required $\alpha\lambda_{\mathrm{CE}}$ assuming an orbit for WD 1856 b such that the primary star was just filling its Roche lobe. We calculated the minimum $\alpha\lambda_{\mathrm{CE}}$ during three different intervals in the progenitor star's evolution: before the star reached the thermally pulsating AGB phase and began rapidly losing mass, before 30% of the progenitor's envelope mass had been lost, and at any point in the star's evolution. Our values for $\alpha\lambda_{\mathrm{CE}}$ as a function of stellar mass and at different points in the progenitor's evolution are also shown in Extended Data Fig. 9. Our curve of the minimum $\alpha\lambda_{\mathrm{CE}}$ before the AGB confirms the results from our analytic study: it is energetically difficult for WD 1856 b to eject the envelope while most of its mass is still in place. Even once 30% of the envelope's mass is lost, it is still difficult to eject the envelope; typical $\alpha\lambda_{\mathrm{CE}}$ values of 1–10 indicate that WD 1856 b's gravitational potential energy is insufficient, but the envelope perhaps could be ejected if a large fraction of the envelope's internal energy contributed to its ejection. By the very end of the AGB phase, once about 50–60% of the envelope's mass has been lost, the minimum $\alpha\lambda_{\mathrm{CE}}$ values become less than unity. The observed population of post-common-envelope binaries suggests[115] that towards the end of the AGB phase, λ_{CE} could be as high as 10, so it is possible that WD 1856 b could eject its progenitor's envelope (though the population also favours values of $\alpha \lesssim 0.3$). However, given the relatively small region of parameter space in which this mechanism could produce the current orbit of WD 1856 b, we consider common-envelope evolution less likely than the dynamical explanation outlined below.

For planets that might manage to eject the envelope of the WD progenitor, at least in principle, there are some other perils that may await it. A previous work[123] examined whether planets and brown dwarfs would be disrupted by ram pressure during their passage through the dense inner envelopes of the giant during the common-envelope phase. They conclude that brown dwarfs and Jovian-mass objects (including a $10M_{\mathrm{J}}$ planet) are not likely to lose a substantial amount of mass during their passage, whereas lower-mass planets could well be destroyed. Another work[124] studied the mass loss of planets that might survive the common envelope, only to find themselves in the intense radiation of the nascent white dwarf[125]. That work[124] concluded that, although lower-mass planets might be obliterated by evaporation, Jovian-mass planets and those of higher mass might well survive to the point where the WD has cooled sufficiently for planetary evaporative losses to become unimportant. Thus, if WD 1856 b had somehow been able to successfully eject the envelope of its progenitor, it might then survive the subsequent heating by the very hot white dwarf. However, we caution that these conclusions are very dependent on the assumed input physics of the models.

Dynamical formation

Given the difficulty in explaining the current orbit of WD 1856 b with common-envelope evolution we investigated other ways to form the system. Here, we consider whether WD 1856 b could have reached its current orbit as a result of dynamical scattering after WD 1856's progenitor evolved into a white dwarf. This framework has two main components: (1) perturbing WD 1856 b into a high-eccentricity orbit with a close periastron passage and (2) dissipating the orbital energy to shrink the planet's semimajor axis and shorten the orbital period to 1.4 days. We consider these two processes separately.

Generating a short periastron distance for WD 1856 b. WD 1856 b must have formed and evolved far away (≳1 AU) from WD 1856's progenitor star, and so we explored whether dynamical processes can perturb a planet with a semimajor axis of roughly 1–2 AU into a highly eccentric orbit with a periastron distance of only a few solar radii. First, we considered whether the gravitational influence of WD 1856 b's M-dwarf companions (G 229-20 A/B) could excite a high eccentricity in WD 1856 b's orbit via the Kozai–Lidov effect[126–128]. We ran a small set of *N*-body simulations using Mercury6[129] with the four known bodies in the WD 1856 system, initialized with WD 1856 b in a circular orbit with a distance of 1–2 AU about WD 1856, and with G 229-20 A/B orbiting at a distance of about 1,000 AU, consistent with the result of our LOFTI orbit fits (described above). Under these conditions (and when the mutual inclination between the orbits of WD 1856 b and G 229-20 A/B is large enough), G 229-20 A/B do induce Kozai–Lidov cycles in WD 1856 b's orbit, but the timescales are slow (≳100 Myr) and the amplitudes of the eccentricity oscillation are low ($e \approx 0.1$). Although different values of initial conditions (including the eccentricities of both orbits) and mutual inclinations may alter the specific amplitudes and timescales of Kozai–Lidov oscillations, we conclude that it is difficult for G 229-20 A/B to excite WD 1856 b's orbit to $e \gtrsim 0.99$ eccentricity and close periastron passages.

Even if G 229-20 A/B could not have decreased WD 1856 b's periastron distance by exciting its eccentricity, it is possible that additional (undiscovered) bodies in the system could have. Previous work[13,25] has shown that systems of multiple planets residing exterior to the red giant expansion radius (but in a relatively well packed configuration) can remain dynamically stable until after the white dwarf has formed and begun cooling, then experience potentially violent instabilities. One of these works[25] found that increasing the number of planets in their simulations resulted in more extreme dynamical evolution, including periastron passages as close as that of WD 1856 b. We ran an additional set of *N*-body simulations to confirm that the pattern seen in ref. [25] holds true for systems with giant planets similar in mass to WD 1856. Again, we used Mercury6 to calculate the evolution of multi-planet systems. We initialized our simulations with up to four planets in closely packed orbits, with equal masses to WD 1856 b. Although our simulations are not an exhaustive exploration of parameter space, they do confirm that in multi-planet systems, violent dynamical instabilities can lead to planets being ejected from the system, sent onto a collision course with the white dwarf, or into orbits with small periastron distances.

Dissipating orbital energy and shrinking the semimajor axis. If WD 1856 b had been perturbed into a highly eccentric orbit with a close periastron passage, it must have dissipated much of its orbital energy to end up with a 1.4-day period as we see today. We investigated whether tidal effects could dissipate WD 1856's orbital energy quickly enough to nearly circularize the planet's orbit in the roughly 5.85-Gyr cooling age of the white dwarf. Because WD 1856 is very small and dense, any tides raised on the white dwarf by the planet will be small and have negligible dissipative effects. Instead, any tidal dissipation in WD 1856 b's orbit must be due to tides raised on the planet by its star.

The problem of tidally dissipating orbital energy for planets in highly eccentric orbits around white dwarfs has previously been studied[27,130]. The authors calculated the total time needed to circularize a highly eccentric orbit as the sum of two different tidal regimes: a chaotic tidal regime at high eccentricities ($e \gtrsim 0.95$), where dissipation is dominated by the exchange in energy between the orbit and internal modes, and a classic tidal regime, at $e \lesssim 0.95$, where dissipation is dominated by equilibrium tides. In the abovementioned works, the authors calculate timescales for the completion of the chaotic tidal regime for gas giant planets and find typical values between 1 and 100 Myr—we conservatively choose a timescale at the high-end of their estimates for the WD 1856 system. We then estimated the time needed for the system to circularize from $e \approx 0.95$ via equilibrium tides with:

$$t_{\mathrm{circ}} = \frac{6a^5 Q_{\mathrm{p}} m_{\mathrm{p}}}{63 n_{\mathrm{p}} k_{\mathrm{p}} m_* R_{\mathrm{p}}^5}, \tag{12}$$

where a is the planetary semimajor axis, Q_{p} is the planetary tidal quality factor, m_{p} the planetary mass, n_{p} the planetary mean motion (related to the orbital period P by $n = 2\pi/P$), k_{p} the planetary Love number, m_* the stellar mass and R_{p} the planetary radius[26]. Plugging in parameters for the WD 1856 system, and assuming WD 1856 b has Jupiter's mass, radius and Q/k_{p} (estimated[131] to be $Q_{\mathrm{J}}/k_{\mathrm{p,J}} \approx 10^5$), we estimate a tidal circularization timescale of about 2 Myr. Larger planet masses ($5M_{\mathrm{J}}$–$10M_{\mathrm{J}}$) and more conservative estimates of Q/k_{p} up to 10^7 should still circularize within the white dwarf's cooling age. All together, the timescale for tidal circularization of WD 1856 b's orbit is comfortably less than the system's age.

We note that these processes could just as easily be applied to smaller planets than WD 1856 b. Packed systems of Earth-mass planets should exhibit the same dynamical instabilities that can drive close periastron distances for giant planets[25], and tidal circularization should be even more efficient for rocky Earth-sized planets than gas giants such as WD 1856. We estimate that tides raised on an Earth-sized planet should dissipate its orbital eccentricity within about 500,000 yr. This formation pathway could potentially lead to the production of habitable-zone rocky planets[29]. Old white dwarfs cool slowly and could provide a relatively stable radiation environment for billions of years[132]; we estimate that WD 1856 b's current orbital location was in the circumstellar habitable zone for almost 3 Gyr. WD 1856 b may demonstrate a mechanism that can lead to a second generation of habitability in a planetary system.

Other theories

We also explored other mechanisms that might have led to WD 1856 b's current orbital configuration. We consider these other mechanisms less likely since they require either finely tuned or a priori unlikely initial conditions to succeed, but mention them for completeness.

Close stellar encounter. WD 1856 may have been perturbed from its initial long-period orbit by a close flyby with another star. We estimated the most likely distance of closest approach D_{closest} between WD 1856 and another star during its 5.85-Gyr cooling age:

$$D_{\mathrm{closest}} \approx (\pi v t_{\mathrm{cool}} n)^{-1/2}, \tag{13}$$

where v is the typical stellar velocity in WD 1856's vicinity (≈60 km s^{-1}), t_{cool} is the cooling age (5.85 Gyr) and n is the number density of stars in the vicinity of WD 1856. We estimated n using the fact that there are about 6,000 stars within 25 pc of the Sun from Gaia DR2, giving a density of about 0.1 star per cubic parsec. We find $D_{\mathrm{closest}} \approx 600$ AU, so probably within its cooling lifetime, another star has passed by within the orbit of G 229-20 A/B. However, a much closer approach than 600 AU would be required to perturb WD 1856 b from an orbit of approximately 1–10 AU to a close periastron passage, and the probability p of such a close approach decreases as $p \propto D_{\mathrm{closest}}^{-2}$.

Dynamical instabilities from galactic tides. A previous study[133] suggested that galactic tides could perturb the orbit of a wide white dwarf binary and lead to a close approach billions of years after the system's formation. This mechanism could provide a trigger for dynamical instabilities in old white dwarf systems. In principle, such a mechanism could be important to the formation of WD 1856 b, given the old system age and the presence of wide visual companions. That work[133] calculated that for galactic tides to be important on timescales of a few Gyr, the semimajor axis must be greater than about a few thousand AU and the wide binary orbit must be highly inclined with respect to the galactic plane (that is, the pole of the orbit must be near the plane). Our fit to the WD 1856/G 229-20 orbit with LOFTI gives a semimajor axis of about 1,500 AU with a tail out beyond 4,000 AU. We constrained the inclination of the orbit with respect to the galactic plane, i_b, by calculating the location of the orbital pole[134,135] for each posterior sample from our fit. In particular, we used the equations on page 13 of ref. [134], after correcting an error in the second equation on page 13 that should read $\sin i \sin \Omega = m \sin M$ (see ref. [136]). The probability distribution for i_b is strongly peaked towards high inclinations, with the greatest probability at 90°. At 68% and 95% confidence, i_b must be greater than 60° and 41°, respectively. Therefore, the galactic tide mechanism could plausibly operate in at least part of the allowed orbital parameter space.

Tidal dissipation during the giant phase. Previous work[137] has calculated the orbital evolution of exoplanets orbiting near expanding giant stars (see also ref. [138]). The orbits of these planets evolve owing to two competing factors: mass loss (which drives orbits outwards) and tidal dissipation (which drives orbits inwards). Planets that orbit near an equilibrium radius where these two effects are nearly equal in strength can in some cases migrate inwards owing to tidal evolution, but avoid engulfment by the red giant host. This requires extremely finely tuned initial parameters to have a chance of reproducing the present-day configuration of WD 1856 b. Computing the exact location of this radius (which is probably somewhere around 1–2 AU) is difficult as the radius depends on the starting angular momentum, the mass-loss rate, the dissipation coefficients and other parameters that are difficult to constrain; however, it is plausible that finely tuning the initial parameters of the planetary orbit and stellar properties could shrink the orbit of WD 1856 b to its current semimajor axis.

Dynamical interactions near periastron. If two planets happened to be scattered into close periastron passages at the same time and had a close scattering event near periastron, one planet could have been ejected, leaving the other planet in a short-period orbit around WD 1856. The likelihood of such an encounter is fairly low; events that can excite high eccentricities and close periastron distances are already rare (happening perhaps once in the lifetime of a white dwarf planetary system)[25], and so the probability of two planets having close periastron passages simultaneously is even lower. Another related mechanism involves a proto-WD 1856 b with a massive moon (or a binary planet) on a highly eccentric orbit with a close periastron passage. The moon/binary companion could be ejected[139] in a similar fashion to hypervelocity stars, which are ejected binary members perturbed by the Galaxy's central black hole[140], shedding enough orbital energy to leave WD 1856 b in a nearly circular orbit. Again, this mechanism is a priori unlikely, because we have yet to discover a binary planet.

Partial tidal disruption. If WD 1856 b reached a periastron distance slightly closer to WD 1856 than the Roche limit, it could have been partially tidally disrupted, losing enough mass to dissipate its orbital energy while remaining at least somewhat intact[141]. This process has also been studied in the case of the tidal disruption of a star by a supermassive black hole[142]. If this process happened recently and material from the planet was still accreting onto the white dwarf, the elements might be visible in the planet's spectrum. This motivates more sensitive spectroscopy of WD 1856.

Expected amplitude of spectral features in transmission

Owing to the small radius of the white dwarf host star, the spectral features expected from transmission spectroscopy are much larger than they would be around a main-sequence star. We estimated the amplitude of spectral features as described below.

Traditionally the amplitude of spectral features in transmission is proportional to the annulus of the planet's terminator region[143]. However, that approximation does not apply to the case of a grazing transit where the star is smaller than the planet. To account for the grazing geometry of WD 1856, we assumed that the atmosphere covers a slice of the star with width equal to the stellar diameter and height equal to the scale height. In this case, the amplitude A of spectral features is

$$A \approx \frac{2nH}{\pi R_*}. \quad (14)$$

Here n is the number of scale heights typically crossed by atmospheric features (usually $n = 2$ for cloud-free gas giant exoplanets[144]) and H is the atmospheric scale height,

$$H = \frac{kT}{\mu g}, \quad (15)$$

where k is Boltzmann's constant, T is the planet's temperature, μ is the mean molecular weight in the atmosphere and g is the planet's surface gravity. To calculate the scale height, we assumed a solar composition atmosphere ($\mu = 2.3$ AMU) and assumed planet properties for two cases: (1) $M_p = 10M_J$, $T = 280$ K (a reasonable internal temperature for an object of this mass); and (2) $M_p = 1M_J$, $T = 165$ K (the equilibrium temperature).

For case 1, the scale height is $H = 4$ km and the amplitude of spectral features is 0.1%. For case 2, the scale height is $H = 12$ km and the amplitude of spectral features is 0.7%.

We note that our assumption that the atmosphere covers a slice of the star with width equal to the stellar diameter is an approximation for nearly 50% deep transits of planets that are much larger than their stars. A more general expression (valid for $|1 - R_p/R_*| < b < 1 + R_p/R_*$) for the expected height of transmission features for grazing transits is

$$A \approx \frac{snH}{\pi R_*}, \quad (16)$$

where

$$s = 2\frac{R_p}{R_*}\cos^{-1}\left(\frac{b^2 - 1 + \left(\frac{R_p}{R_*}\right)^2}{2b\frac{R_p}{R_*}}\right). \quad (17)$$

For cases similar to WD 1856—where the planet is much larger than the star and blocks close to 50% of the stellar disk—$s \approx 2$, and the expression reduces to equation (14). For WD 1856 b's particular transit parameters, $s = 2.004$.

Expected amplitude of Doppler-boosting signal

WD 1856 b's mass could be measurable via small variations in the host star's brightness caused by Doppler boosting[145,146]. The semi-amplitude A_b of the Doppler boosting signal is

$$A_b = (3 - \alpha)\frac{K}{c}, \quad (18)$$

where K is radial velocity semi-amplitude induced by the planet, c is the speed of light and α is the average logarithmic derivative of flux with respect to frequency. For a blackbody spectrum, α is approximately

$$\alpha \approx 3 - \frac{xe^x}{e^x - 1}, \quad (19)$$

and

$$x = \frac{h\nu}{kT_{\text{eff}}} \quad (20)$$

where h is Planck's constant, v is the frequency of light in the observed bandpass, k is Boltzmann's constant, e is Euler's number and T_{eff} is the blackbody temperature. Assuming a mass of $14M_{\mathrm{J}}$ for WD 1856 b, the Doppler boosting amplitude is about 50 ppm in the TESS bandpass, about 100 ppm in blue optical light, and about 30 ppm in near-infrared light around 1.5 μm.

It will be difficult to detect these signals because of WD 1856's intrinsic faintness and contamination from G 229-20 A/B. We fitted the out-of-transit TESS light curve (with a dilution correction applied) with a sine/cosine model and found a boosting semi-amplitude of −770 ± 1,130 ppm—far too uncertain to detect an orbiting planet. If the PLATO mission[147] observes WD 1856 near the centre of its field of view for two years, it may come close to a tentative detection of a $14M_{\mathrm{J}}$ planet, depending on how much starlight from G 229-20 A/B contaminates the aperture of WD 1856. With their large apertures and high spatial resolution, JWST and the Hubble Space Telescope could detect the boosting signal, but the observations would be expensive—that is, time consuming. A 3σ detection of a $14M_{\mathrm{J}}$ planet would probably require ≥10 days of observations.

Data availability

We provide all reduced light curves and spectra with the manuscript. The Spitzer images are available for download at the Spitzer Heritage Archive (http://irsa.ipac.caltech.edu/applications/Spitzer/SHA/), and the TESS images and light curves are available from the Mikulski Archive for Space Telescopes (https://archive.stsci.edu/tess/). Source data are provided with this paper.

Code availability

Much of the code used to produce these results is publicly available and linked throughout the paper. We wrote custom software to analyse the data collected in this project. Though this code was not written with distribution in mind, it is available online at https://github.com/avanderburg/.

31. Ricker, G. R. et al. Transiting Exoplanet Survey Satellite (TESS). *J. Astron. Telesc. Instrum. Syst.* **1**, 014003 (2014).
32. Dufour, P. et al. The Montreal White Dwarf Database: a tool for the community. In *20th European White Dwarf Workshop (EuroWD16)* (eds Tremblay, P.-E., Gaensicke, B. & Marsh, T.) 3–8 (2017).
33. Stassun. K. G. et al. The TESS Input Catalog and candidate target list. *Astron. J.* **156**, 102 (2018); correction **156**, 183 (2018).
34. Gould, A. & Morgan, C. W. Transit target selection using reduced proper motions. *Astrophys. J.* **585**, 1056–1061 (2003).
35. Altmann, M., Roeser, S., Demleitner, M., Bastian, U. & Schilbach, E. Hot Stuff for One Year (HSOY). A 583 million star proper motion catalogue derived from Gaia DR1 and PPMXL. *Astron. Astrophys.* **600**, L4 (2017).
36. Gentile Fusillo, N. P. et al. A Gaia Data Release 2 catalogue of white dwarfs and a comparison with SDSS. *Mon. Not. R. Astron. Soc.* **482**, 4570–4591 (2019).
37. Jenkins, J. M. Overview of the TESS Science Pipeline. In *AAS/Division for Extreme Solar Systems III* (chairs Mayor, M. & Rasio, F.) 106.05 (2015).
38. Jenkins, J. M. et al. The TESS science processing operations center. In *Proc. SPIE 9913 Software and Cyberinfrastructure for Astronomy IV* (eds Chiozzi, G. & Guzman, J. C.) 99133E (2016).
39. Smith, J. C. et al. Kepler presearch data conditioning II—a Bayesian approach to systematic error correction. *Publ. Astron. Soc. Pacif.* **124**, 1000–1014 (2012).
40. Stumpe, M. C. et al. Multiscale systematic error correction via wavelet-based bandsplitting in Kepler data. *Publ. Astron. Soc. Pacif.* **126**, 100 (2014).
41. Jenkins, J. M. The impact of solar-like variability on the detectability of transiting terrestrial planets. *Astrophys. J.* **575**, 493–505 (2002).
42. Evans, D. F. Evidence for unresolved exoplanet-hosting binaries in Gaia DR2. *Res. Notes AAS* **2**, 20 (2018).
43. Rizzuto, A. C. et al. Zodiacal Exoplanets in Time (ZEIT). VIII. A two-planet system in Praesepe from K2 Campaign 16. *Astron. J.* **156**, 195 (2018).
44. Lindegren, L. *Re-normalising the Astrometric Chi-Square in Gaia DR2* Gaia Technical Note No. GAIA-C3-TN-LU-LL-124-01 (Gaia DPAC, 2018).
45. Abell, G. O. Globular clusters and planetary nebulae discovered on the National Geographic Society–Palomar Observatory Sky Survey. *Publ. Astron. Soc. Pacif.* **67**, 258–261 (1955).
46. Rappaport, S. et al. Drifting asteroid fragments around WD 1145+017. *Mon. Not. R. Astron. Soc.* **458**, 3904–3917 (2016).
47. Narita, N. et al. MuSCAT2: four-color simultaneous camera for the 1.52-m Telescopio Carlos Sánchez. *J. Astron. Telesc. Instrum. Syst.* **5**, 015001 (2019).
48. Schmidt, G. D., Weymann, R. J. & Foltz, C. B. A. Moderate-resolution, high-throughput CCD channel for the MMT Spectrograph. *Publ. Astron. Soc. Pacif.* **101**, 713 (1989).
49. Miller, J. S. & Stone, R. P. *The Kast Double Spectograph* Lick Observatory Technical Report 66 (University of California Observatories/Lick Observatory, 1994).
50. Chonis, T. S., Hill, G. J., Lee, H., Tuttle, S. E. & Vattiat, B. L. LRS2: the new facility low resolution integral field spectrograph for the Hobby–Eberly telescope. In *Proc. SPIE Astronomical Telescopes and Instrumentation* Vol. 9147 (eds Ramsay, S. K., McLean, I. S. & Takami, H.) 91470A (SPIE, 2014).
51. Zeimann, G. *Panacea* source code (accessed 24 June 2020); https://github.com/grzeimann/Panacea (2019).
52. Elias, J. H. et al. Design of the Gemini near-infrared spectrograph. In *Proc. Ground-based and Airborne Instrumentation for Astronomy* (eds McLean, I. S. & Iye, M.) 62694C (2006).
53. Mason, R. E. et al. The nuclear near-infrared spectral properties of nearby galaxies. *Astrophys. J. Suppl. Ser.* **217**, 13 (2015).
54. Telting, J. H. et al. FIES: the high-resolution Fiber-fed Echelle Spectrograph at the Nordic Optical Telescope. *Astron. Nachr.* **335**, 41 (2014).
55. Stempels, E. & Telting, J. FIEStool: automated data reduction for FIber-fed Echelle Spectrograph (FIES) *Astrophysics Source Code Library* http://ascl.net/1708.009 (2017).
56. Fűrész, G. Design and Application of High Resolution and Multiobject Spectrographs: Dynamical Studies of Open Clusters. PhD thesis, Univ. Szeged (2008).
57. Buchhave, L. A. et al. An abundance of small exoplanets around stars with a wide range of metallicities. *Nature* **486**, 375–377 (2012).
58. Stefanik, R. P., Latham, D. W. & Torres, G. Radial-velocity standard stars. In *IAU Colloquium 170: Precise Stellar Radial Velocities* Vol. 185 (eds Hearnshaw, J. B. & Scarfe, C. D.) 354–366 (1999).
59. Lépine, S. et al. A spectroscopic catalog of the brightest (J < 9) M dwarfs in the northern sky. *Astron. J.* **145**, 102 (2013).
60. Cubillos, P. et al. WASP-8b: characterization of a cool and eccentric exoplanet with Spitzer. *Astrophys. J.* **768**, 42 (2013).
61. Xu, S. & Jura, M. Spitzer observations of white dwarfs: the missing planetary debris around DZ stars. *Astrophys. J.* **745**, 88 (2012).
62. Xu, S. et al. Infrared variability of two dusty white dwarfs. *Astrophys. J.* **866**, 108 (2018).
63. Blouin, S., Dufour, P., Thibeault, C. & Allard, N. F. A new generation of cool white dwarf atmosphere models. IV. Revisiting the spectral evolution of cool white dwarfs. *Astrophys. J.* **878**, 63 (2019).
64. Blouin, S., Dufour, P. & Allard, N. F. A new generation of cool white dwarf atmosphere models. I. Theoretical framework and applications to DZ stars. *Astrophys. J.* **863**, 184 (2018).
65. Kowalski, P. M. Infrared absorption of dense helium and its importance in the atmospheres of cool white dwarfs. *Astron. Astrophys.* **566**, L8 (2014).
66. Stassun, K. G., Corsaro, E., Pepper, J. A. & Gaudi, B. S. Empirical accurate masses and radii of single stars with TESS and Gaia. *Astron. J.* **155**, 22 (2018).
67. Eggleton, P. *Evolutionary Processes in Binary and Multiple Stars* (Cambridge Univ. Press, 2006).
68. Zapolsky, H. S. & Salpeter, E. E. The mass–radius relation for cold spheres of low mass. *Astrophys. J.* **158**, 809 (1969).
69. Mestel, L. On the theory of white dwarf stars. I. The energy sources of white dwarfs. *Mon. Not. R. Astron. Soc.* **112**, 583 (1952).
70. van Horn, H. M. Cooling of white dwarfs. In *International Astronomical Union Symp. No. 42: White Dwarfs* (ed. Luyten, W. J.) 97–115 (Reidel, 1971).
71. Mann, A. W., Feiden, G. A., Gaidos, E., Boyajian, T. & von Braun, K. How to constrain your M dwarf: measuring effective temperature, bolometric luminosity, mass, and radius. *Astrophys. J.* **804**, 64 (2015); erratum **819**, 87 (2016).
72. Mann, A. W. et al. How to constrain your M dwarf. II. The mass–luminosity–metallicity relation from 0.075 to 0.70 Solar masses. *Astrophys. J.* **871**, 63 (2019).
73. Stassun, K. G. et al. The revised TESS input catalog and candidate target list. *Astron. J.* **158**, 138 (2019).
74. Pearce, L. A. *Linear Orbits for the Impatient* (accessed 24 June 2020); https://github.com/logan-pearce/LOFTI (2019).
75. Pearce, L. A. et al. Orbital parameter determination for wide stellar binary systems in the age of Gaia. *Astrophys. J.* **894**, 115 (2020).
76. Blunt, S. et al. Orbits for the Impatient: a Bayesian rejection-sampling method for quickly fitting the orbits of long-period exoplanets. *Astron. J.* **153**, 229 (2017).
77. Eastman, J., Siverd, R. & Gaudi, B. S. Achieving better than 1 minute accuracy in the heliocentric and barycentric Julian dates. *Publ. Astron. Soc. Pacif.* **122**, 935 (2010).
78. Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J. Lett.* **580**, 171–175 (2002).
79. Eastman, J., Gaudi, B. S. & Agol, E. EXOFAST: a fast exoplanetary fitting suite in IDL. *Publ. Astron. Soc. Pacif.* **125**, 83–112 (2013).
80. Gianninas, A., Strickland, B. D., Kilic, M. & Bergeron, P. Limb-darkening coefficients for eclipsing white dwarfs. *Astrophys. J.* **766**, 3 (2013).
81. Claret, A. et al. Gravity and limb-darkening coefficients for compact stars: DA, DB, and DBA eclipsing white dwarfs. *Astron. Astrophys.* **634**, A93 (2020).
82. Claret, A. & Bloemen, S. Gravity and limb-darkening coefficients for the Kepler, CoRoT, Spitzer, uvby, UBVRIJHK, and Sloan photometric systems. *Astron. Astrophys.* **529**, A75 (2011).
83. Seager, S. & Mallén-Ornelas, G. A unique solution of planet and star parameters from an extrasolar planet transit light curve. *Astrophys. J.* **585**, 1038–1055 (2003).
84. Lucy, L. B. & Sweeney, M. A. Spectroscopic binaries with circular orbits. *Astron. J.* **76**, 544–556 (1971).
85. Goodman, J. & Weare, J. Ensemble samplers with affine invariance. *Comm. App. Math. Comp. Sci.* **5**, 65–80 (2010).
86. Kopal, Z. *Close Binary Systems* (Chapman & Hall, 1959).

87. Kipping, D. M. Efficient, uninformative sampling of limb darkening coefficients for two-parameter laws. *Mon. Not. R. Astron. Soc.* **435**, 2152–2160 (2013).
88. Saumon, D. & Marley, M. S. The evolution of L and T dwarfs in color–magnitude diagrams. *Astrophys. J.* **689**, 1327–1344 (2008).
89. Nelson, L. A., Rappaport, S. A. & Joss, P. C. On the nature of the companion to Van Biesbroeck 8. *Nature* **316**, 42–44 (1985).
90. Chabrier, G., Johansen, A., Janson, M. & Rafikov, R. Giant planet and brown dwarf formation. In *Protostars and Planets VI* (eds Beuther, H. et al.) 619–642 (Univ. Arizona Press, 2014).
91. Bowler, B. P., Blunt, S. C. & Nielsen, E. L. Population-level eccentricity distributions of imaged exoplanets and brown dwarf companions: dynamical evidence for distinct formation channels. *Astron. J.* **159**, 63 (2020).
92. Phillips, M. W. et al. A new set of atmosphere and evolution models for cool T–Y brown dwarfs and giant exoplanets. *Astron. Astrophys.* **637**, A38 (2020).
93. Miles, B. E. et al. Observations of disequilibrium CO chemistry in the coldest brown dwarfs. *Astron. J.* **160**, 63 (2020).
94. Morley, C. V. et al. An L band spectrum of the coldest brown dwarf. *Astrophys. J.* **858**, 97 (2018).
95. Morley, C. V. et al. Water clouds in Y dwarfs and exoplanets. *Astrophys. J.* **787**, 78 (2014).
96. Shappee, B. J. et al. The man behind the curtain: X-rays drive the UV through NIR variability in the 2013 active galactic nucleus outburst in NGC 2617. *Astrophys. J.* **788**, 48 (2014).
97. Kochanek, C. S. et al. The All-Sky Automated Survey for Supernovae (ASAS-SN) Light Curve Server v1.0. *Publ. Astron. Soc. Pacif.* **129**, 104502 (2017).
98. Butters, O. W. et al. The first WASP public data release. *Astron. Astrophys.* **520**, L10 (2010).
99. Gizis, J. E. M-subdwarfs: spectroscopic classification and the metallicity scale. *Astron. J.* **113**, 806–822 (1997).
100. Lépine, S., Rich, R. M. & Shara, M. M. Revised metallicity classes for low-mass stars: dwarfs (dM), subdwarfs (sdM), extreme subdwarfs (esdM), and ultrasubdwarfs (usdM). *Astrophys. J.* **669**, 1235–1247 (2007).
101. Mann, A. W., Brewer, J. M., Gaidos, E., Lépine, S. & Hilton, E. J. Prospecting in late-type dwarfs: a calibration of infrared and visible spectroscopic metallicities of late K and M dwarfs spanning 1.5 dex. *Astron. J.* **145**, 52 (2013).
102. Newton, E. R. et al. The Hα emission of nearby M dwarfs and its relation to stellar rotation. *Astrophys. J.* **834**, 85 (2017).
103. West, A. A. et al. The Sloan Digital Sky Survey data release 7 spectroscopic M dwarf catalog. I. Data. *Astron. J.* **141**, 97 (2011).
104. Coşkunoğlu, B. et al. Local stellar kinematics from RAVE data—I. Local standard of rest. *Mon. Not. R. Astron. Soc.* **412**, 1237–1245 (2011).
105. Bensby, T., Feltzing, S. & Oey, M. S. Exploring the Milky Way stellar disk. A detailed elemental abundance study of 714 F and G dwarf stars in the solar neighbourhood. *Astron. Astrophys.* **562**, A71 (2014).
106. Carrillo, A., Hawkins, K., Bowler, B. P., Cochran, W. & Vanderburg, A. Know thy star, know thy planet: chemo-kinematically characterizing TESS targets. *Mon. Not. R. Astron. Soc.* **491**, 4365–4381 (2020).
107. Kilic, M. et al. The ages of the thin disk, thick disk, and the halo from nearby white dwarfs. *Astrophys. J.* **837**, 162 (2017).
108. Haywood, M., Di Matteo, P., Lehnert, M. D., Katz, D. & Gómez, A. The age structure of stellar populations in the solar vicinity. Clues of a two-phase formation history of the Milky Way disk. *Astron. Astrophys.* **560**, A109 (2013).
109. Xiang, M. et al. The ages and masses of a million Galactic-disk main-sequence turnoff and subgiant stars from the LAMOST Galactic Spectroscopic Surveys. *Astrophys. J. Suppl. Ser.* **232**, 2 (2017).
110. Sharma, S. et al. The K2-HERMES Survey: age and metallicity of the thick disc. *Mon. Not. R. Astron. Soc.* **490**, 5335–5352 (2019).
111. Webbink, R. F. Double white dwarfs as progenitors of R Coronae Borealis stars and type I supernovae. *Astrophys. J.* **277**, 355–360 (1984).
112. Pfahl, E., Rappaport, S. & Podsiadlowski, P. The Galactic population of low- and intermediate-mass X-ray binaries. *Astrophys. J.* **597**, 1036–1048 (2003).
113. Zorotovic, M., Schreiber, M. R., Gänsicke, B. T. & Nebot Gómez-Morán, A. Post-common-envelope binaries from SDSS. IX: Constraining the common-envelope efficiency. *Astron. Astrophys.* **520**, A86 (2010).
114. De Marco, O. et al. On the α formalism for the common envelope interaction. *Mon. Not. R. Astron. Soc.* **411**, 2277–2292 (2011).
115. Camacho, J. et al. Monte Carlo simulations of post-common-envelope white dwarf + main sequence binaries: comparison with the SDSS DR7 observed sample. *Astron. Astrophys.* **566**, A86 (2014).
116. Taam, R. E., Bodenheimer, P. & Ostriker, J. P. Double core evolution. I. A 16 $M_\odot$ star with a 1 $M_\odot$ neutron-star companion. *Astrophys. J.* **222**, 269–280 (1978).
117. Taam, R. E. & Bodenheimer, P. The common envelope evolution of massive stars. In *X-Ray Binaries and Recycled Pulsars: Proc. NATO Advanced Research Workshop on X-Ray Binaries and the Formation of Binary and Millisecond Radio Pulsars* (eds van den Heuvel, E. P. & Rappaport, S. A.) 281–291 (Springer Dordrecht, 1992).
118. Tauris, T. M. & Dewi, J. D. M. On the binding energy parameter of common envelope evolution. Dependency on the definition of the stellar core boundary during spiral-in. *Astron. Astrophys.* **369**, 170–173 (2001).
119. Rappaport, S. et al. Discovery of two new thermally bloated low-mass white dwarfs among the Kepler binaries. *Astrophys. J.* **803**, 82 (2015).
120. Choi, J. et al. Mesa Isochrones and Stellar Tracks (MIST). I. Solar-scaled models. *Astrophys. J.* **823**, 102 (2016).
121. Rappaport, S., Podsiadlowski, P., Joss, P. C., Di Stefano, R. & Han, Z. The relation between white dwarf mass and orbital period in wide binary radio pulsars. *Mon. Not. R. Astron. Soc.* **273**, 731–741 (1995).
122. Kalomeni, B. et al. Evolution of cataclysmic variables and related binaries containing a white dwarf. *Astrophys. J.* **833**, 83 (2016).
123. Passy, J.-C., Mac Low, M.-M. & De Marco, O. On the survival of brown dwarfs and planets engulfed by their giant host star. *Astrophys. J. Lett.* **759**, 30 (2012).
124. Bear, E. & Soker, N. Evaporation of Jupiter-like planets orbiting extreme horizontal branch stars. *Mon. Not. R. Astron. Soc.* **414**, 1788–1792 (2011).
125. Schreiber, M. R., Gänsicke, B. T., Toloza, O., Hernandez, M.-S. & Lagos, F. Cold giant planets evaporated by hot white dwarfs. *Astrophys. J.* **887**, L4 (2019).
126. Kozai, Y. Secular perturbations of asteroids with high inclination and eccentricity. *Astron. J.* **67**, 591–598 (1962).
127. Lidov, M. L. The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies. *Planet. Space Sci.* **9**, 719–759 (1962).
128. Stephan, A. P., Naoz, S. & Zuckerman, B. Throwing icebergs at white dwarfs. *Astrophys. J. Lett.* **844**, 16 (2017).
129. Chambers, J. E. A hybrid symplectic integrator that permits close encounters between massive bodies. *Mon. Not. R. Astron. Soc.* **304**, 793–799 (1999).
130. Veras, D. & Fuller, J. The dynamical history of the evaporating or disrupted ice giant planet around white dwarf WD J0914+1914. *Mon. Not. R. Astron. Soc.* **492**, 6059–6066 (2019).
131. Lainey, V., Arlot, J.-E., Karatekin, Ö. & van Hoolst, T. Strong tidal dissipation in Io and Jupiter from astrometric observations. *Nature* **459**, 957–959 (2009).
132. Kozakis, T., Kaltenegger, L. & Hoard, D. W. UV surface environments and atmospheres of Earth-like planets orbiting white dwarfs. *Astrophys. J.* **862**, 69 (2018).
133. Bonsor, A. & Veras, D. A wide binary trigger for white dwarf pollution. *Mon. Not. R. Astron. Soc.* **454**, 53–63 (2015).
134. Chang, Y. C. A study of the orientation of the orbit-planes of 16 visual binaries having determinate inclinations. *Astron. J.* **40**, 11–15 (1929).
135. Agati, J. L. et al. Are the orbital poles of binary stars in the solar neighbourhood anisotropically distributed? *Astron. Astrophys.* **574**, A6 (2015).
136. Heintz, W. D. A statistical study of binary stars. *J. Roy. Astron. Soc. Can.* **63**, 275 (1969).
137. Adams, F. C. & Bloch, A. M. Evolution of planetary orbits with stellar mass loss and tidal dissipation. *Astrophys. J.* **777**, L30 (2013).
138. Rasio, F. A., Tout, C. A., Lubow, S. H. & Livio, M. Tidal decay of close planetary orbits. *Astrophys. J.* **470**, 1187 (1996).
139. Payne, M. J., Veras, D., Holman, M. J. & Gänsicke, B. T. Liberating exomoons in white dwarf planetary systems. *Mon. Not. R. Astron. Soc.* **457**, 217–231 (2016).
140. Bromley, B. C., Kenyon, S. J., Geller, M. J. & Brown, W. R. Binary disruption by massive black holes: hypervelocity stars, S stars, and tidal disruption events. *Astrophys. J.* **749**, L42 (2012).
141. Faber, J. A., Rasio, F. A. & Willems, B. Tidal interactions and disruptions of giant planets on highly eccentric orbits. *Icarus* **175**, 248–262 (2005).
142. Mainetti, D. et al. The fine line between total and partial tidal disruption events. *Astron. Astrophys.* **600**, A124 (2017).
143. Kreidberg, L. Exoplanet atmosphere measurements from transmission spectroscopy and other planet star combined light observations. In *Handbook of Exoplanets* (eds Deeg, H. J. & Belmonte, J. A.) 2083–2105 (2018).
144. Stevenson, K. B. Quantifying and predicting the presence of clouds in exoplanet atmospheres. *Astrophys. J.* **817**, L16 (2016).
145. Loeb, A. & Gaudi, B. S. Periodic flux variability of stars due to the reflex Doppler effect induced by planetary companions. *Astrophys. J. Lett.* **588**, 117–120 (2003).
146. van Kerkwijk, M. H. et al. Observations of Doppler boosting in Kepler light curves. *Astrophys. J.* **715**, 51–58 (2010).
147. Rauer, H. et al. The PLATO 2.0 mission. *Exp. Astron.* **38**, 249–330 (2014).
148. Chambers, K. C. et al. The Pan-STARRS1 surveys. Preprint at: https://www.arxiv.org/abs/1612.05560 (2016).
149. Skrutskie, M. F. et al. The Two Micron All Sky Survey (2MASS). *Astron. J.* **131**, 1163–1183 (2006).
150. Cutri, R. M. et al. VizieR Online Data Catalog: AllWISE Data Release (Cutri+ 2013). *VizieR Online Data Catalog* II/328 (accessed 5 October 2019); http://vizier.u-strasbg.fr/viz-bin/VizieR?-source=II/328

Acknowledgements We thank S. Lepine for providing the archival spectrum of G 229-20 A, and P. Berlind and J. Irwin for collecting and extracting velocities from the TRES spectrum. We thank B.-O. Demory for comments on the manuscript, and F. Rasio, D. Veras, P. Gao, B. Kaiser, W. Torres, J. Irwin, J. J. Hermes, J. Eastman, A. Shporer and K. Hawkins for conversations. A.V.'s work was performed under contract with the California Institute of Technology (Caltech)/Jet Propulsion Laboratory (JPL) funded by NASA through the Sagan Fellowship Program executed by the NASA Exoplanet Science Institute. I.J.M.C. acknowledges support from the NSF through grant AST-1824644, and from NASA through Caltech/JPL grant RSA-1610091. T.D. acknowledges support from MIT's Kavli Institute as a Kavli postdoctoral fellow. D.D. acknowledges support from NASA through Caltech/JPL grant RSA-1006130 and through the TESS Guest Investigator programme, grant 80NSSC19K1727. S.B. acknowledges support from the Laboratory Directed Research and Development programme of Los Alamos National Laboratory under project number 20190624PRD2. C.M. and B.Z. acknowledge support from NSF grants SPG-1826583 and SPG-1826550. A.V. was a NASA Sagan Fellow; J.C.B. is a 51 Pegasi b Fellow; L.A.P. is an NSF Graduate Research Fellow; A.C. is a Large Synoptic Survey Telescope Corporation Data Science Fellow; T.D. is a Kavli Fellow; and C.X.H. is a Juan Carlos Torres Fellow. Resources supporting this work were provided by the NASA High-End Computing (HEC) programme through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center for the production of the SPOC data products. This work is partially based on observations made with the Nordic Optical Telescope, operated by the Nordic Optical Telescope Scientific Association at the Observatorio del Roque de los Muchachos, La Palma, Spain, of the Instituto de Astrofisica de Canarias. This article is partly based on observations made with the MuSCAT2 instrument, developed by ABC, at Telescopio Carlos Sánchez operated on the island of Tenerife by the IAC in the Spanish Observatorio del Teide.
This work is partly supported by JSPS KAKENHI, grant numbers JP17H04574, JP18H01265 and JP18H05439, and JST PRESTO grant number JPMJPR1775. This research has made use of NASA's Astrophysics Data System, the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program, and the SIMBAD database, operated

at CDS, Strasbourg, France. This work is based in part on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. This work is partially based on observations obtained at the International Gemini Observatory, a program of NOIRLab, which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation, on behalf of the Gemini Observatory partnership: the National Science Foundation (United States), National Research Council (Canada), Agencia Nacional de Investigación y Desarrollo (Chile), Ministerio de Ciencia, Tecnología e Innovación (Argentina), Ministério da Ciência, Tecnologia, Inovações e Comunicações (Brazil), and Korea Astronomy and Space Science Institute (Republic of Korea). The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the Indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

Author contributions A.V. led the TESS proposals, identified the planet candidate, organized observations, performed the transit and flux limit analysis, and wrote the majority of the manuscript. S.A.R. helped to organize observations, performed independent data analysis, and wrote portions of the manuscript. S.X. helped to organize observations, obtained and analysed the Gemini data, measured fluxes from the Spitzer data, and helped to guide the strategy of the manuscript. I.J.M.C., L. Kreidberg, V.G., B.B., D.B., J.L.C., D.D., C.D., X.G., S.R.K., F. Morales and L.Y. acquired and produced a light curve from the Spitzer data. S.A.R., J.C.B., L.N., B.Z., F.C.A. and J.J.L. investigated the formation of the WD 1856 system. B.G., F. Murgas, T.G.K., E.P., H.P., A.F. and N.N. acquired follow-up photometry. S.B., P.D. and K.G.S. determined the parameters of the white dwarf, and A.W.M. and E.R.N. studied the M-dwarf companions. C.M., G.Z., W.R.B., R.T., B.K., L.A.B., A.E.D. and A.I.H. acquired spectra of the white dwarf and/or M-dwarf companions. B.M.M., K.H. and T.D. performed an independent analysis of the TESS data, and J.A.L. performed an independent analysis of the white dwarf SED. C.V.M. provided expertise on brown dwarf models, and L. Kaltenegger investigated the system's implications. L.A.P. determined parameters for the binary M-dwarf orbits and white dwarf/M-dwarf orbits, A.C. investigated the system's galactic kinematics. G.R.R., R.K.V., D.W.L., S.S., J.N.W., J.M.J., D.A.C., K.A.C., K.D.C., J.D., A.G., N.M.G., C.X.H., J.P., M.E.R. and J.C.S. are members of the TESS mission team.

Competing interests The authors declare no competing interests.

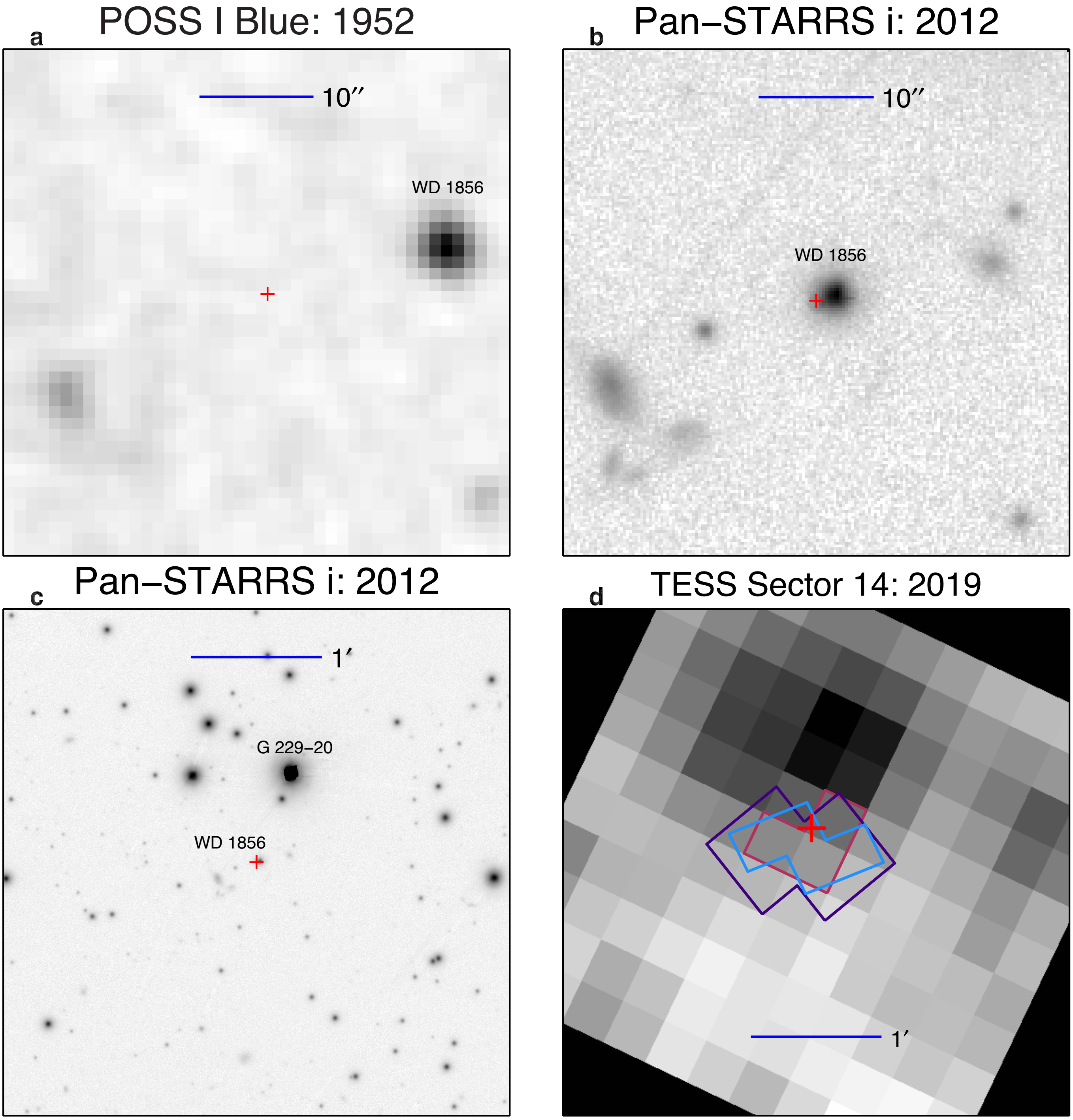
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2713-y.

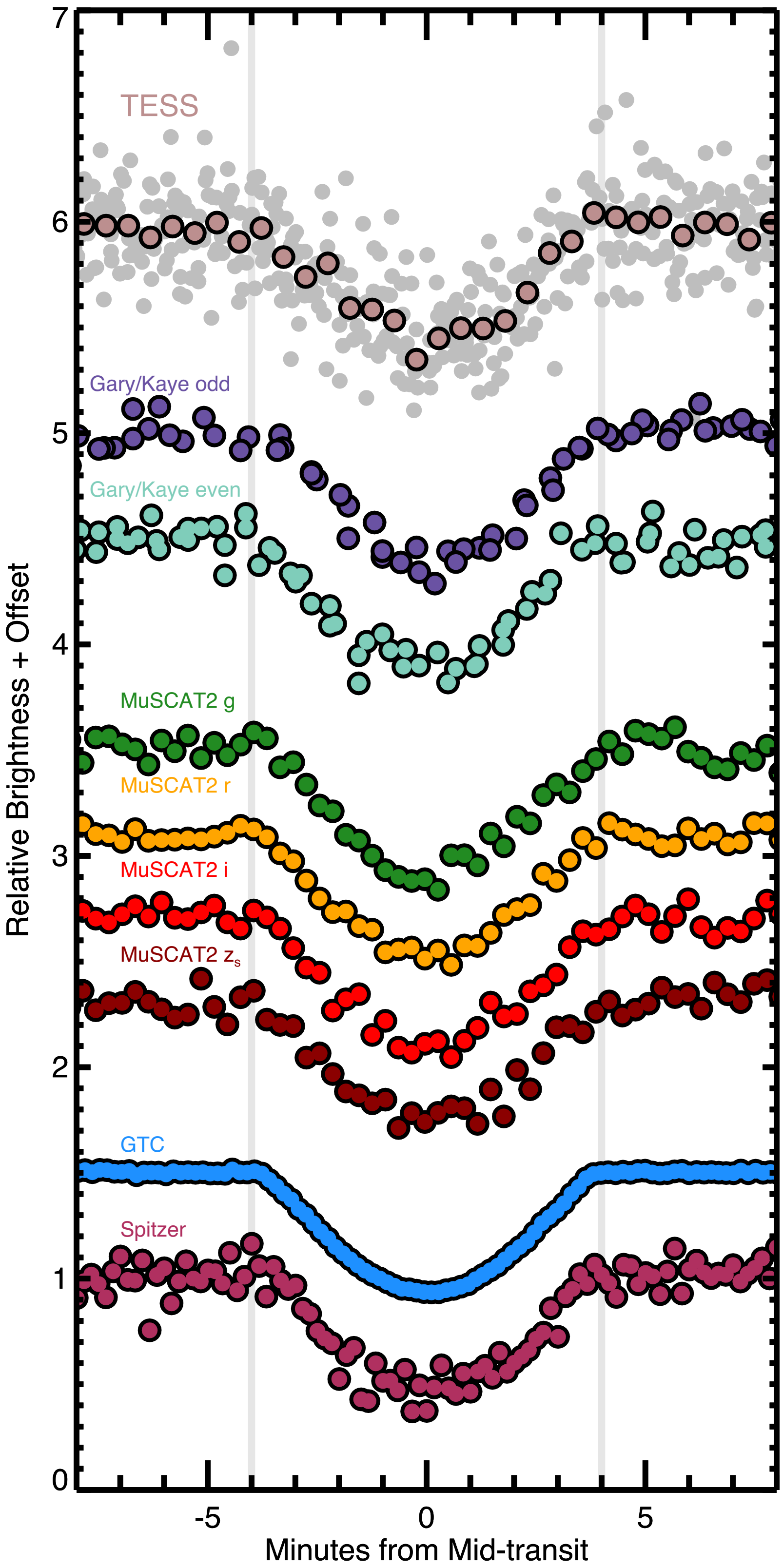
Correspondence and requests for materials should be addressed to A.V.

Peer review information *Nature* thanks Artie Hatzes, Steven Parsons and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

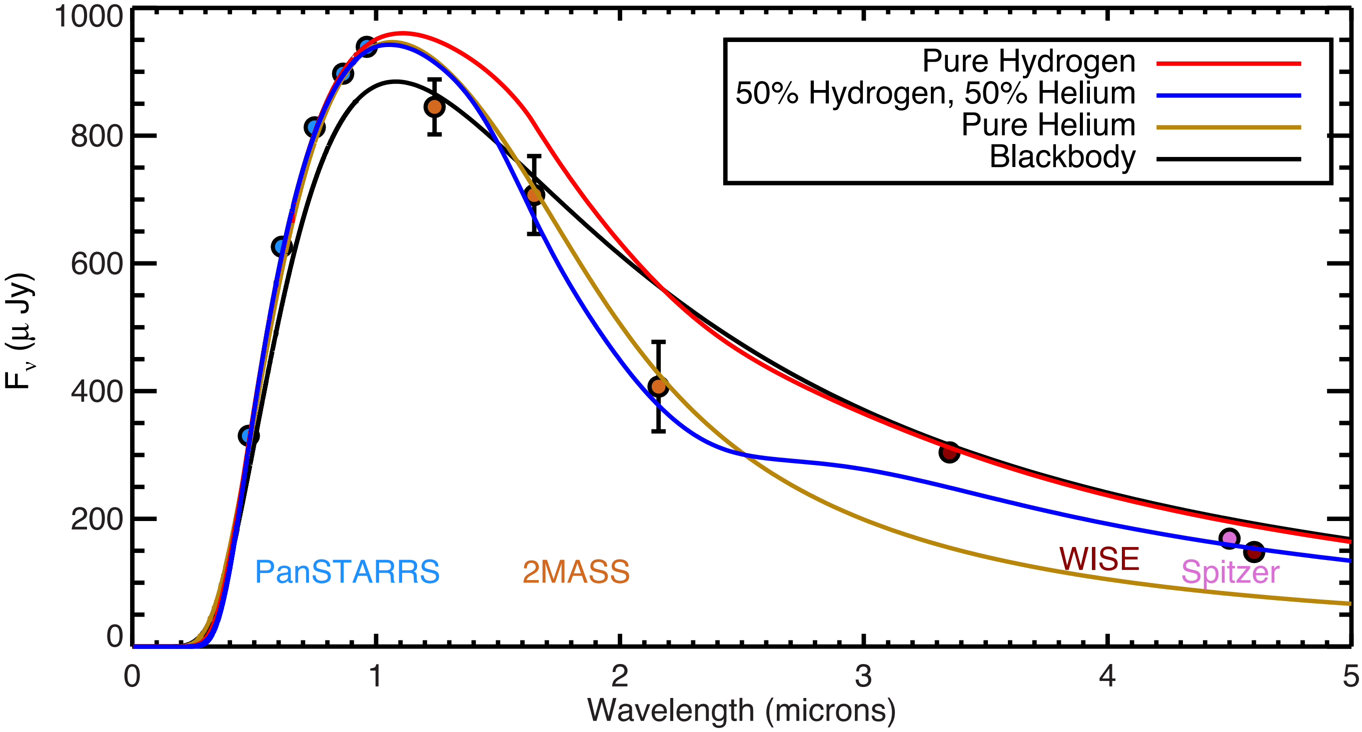
Reprints and permissions information is available at http://www.nature.com/reprints.



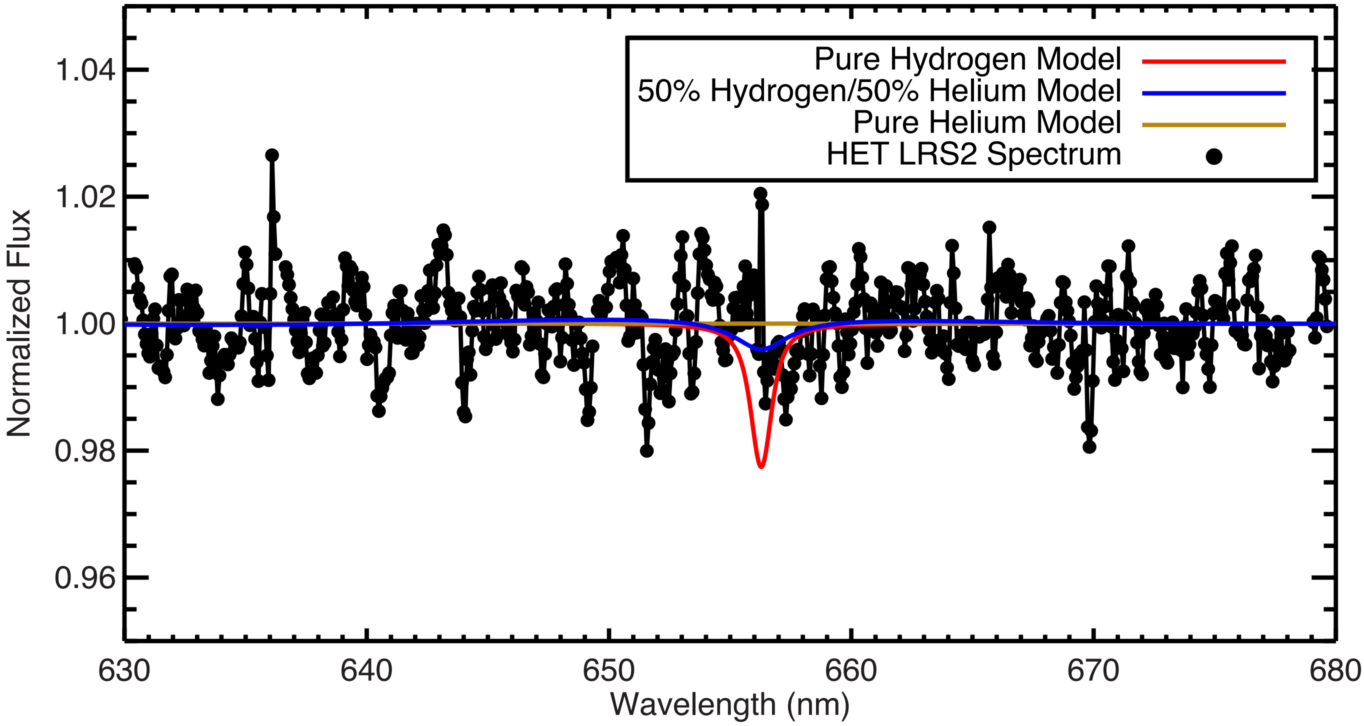
Extended Data Fig. 1 | Archival imaging of WD 1856. a, From the Palomar Observatory Sky Survey on a photographic plate with a blue-sensitive emulsion. **b**, From the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) survey in the i band. **c**, From the Pan-STARRS survey in the i band, zoomed out to show the co-moving M-dwarf pair (labelled G 229-20). **d**, Coadded TESS image from sector 14. The photometric apertures for the three sectors of TESS observations (14, 15 and 19) are shown as red-, purple- and blue-coloured outlines, respectively. The present-day location of WD 1856 is shown with a red cross in all images.



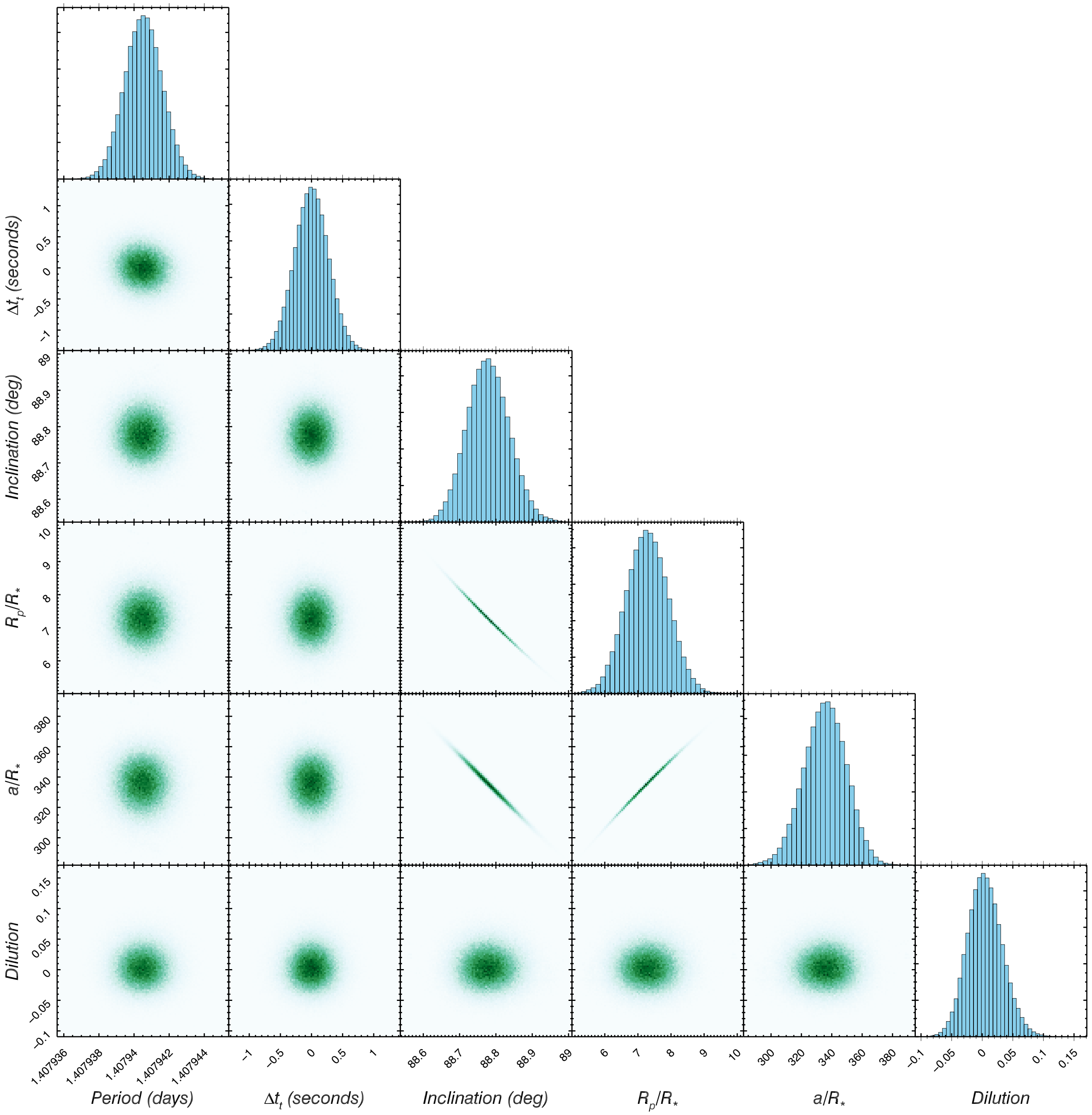
Extended Data Fig. 2 | All transit observations of WD 1856. From top to bottom, we show the light curves (arbitrarily offset for visual clarity) from TESS; data from several private telescopes in Arizona (operated by B.G. and T.G.K.) with odd and even-numbered transits shown separately; simultaneous light curves in four colours from MuSCAT2; a light curve from the GTC, and a light curve from Spitzer. The individual two-minute-cadence TESS flux measurements are shown as grey points, and the rose-coloured points are averages of the brightness in roughly 30 s in orbital phase. The TESS data have been corrected for dilution from nearby stars so that the transit depth matches that of the GTC data.



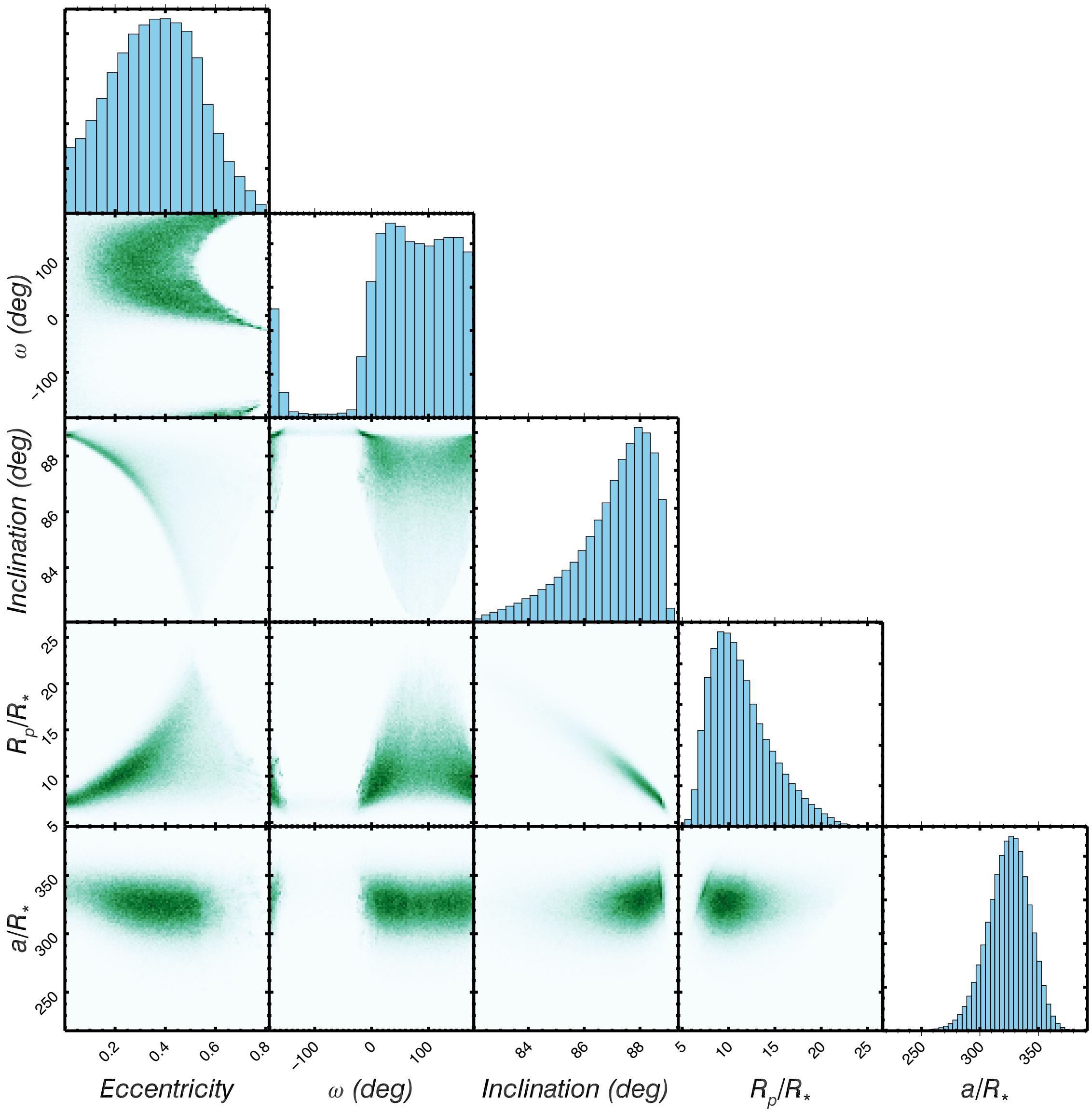
Extended Data Fig. 3 | Spectral energy distribution of WD 1856. Photometric measurements from Pan-STARRS[148], 2MASS[149], WISE[150] and Spitzer are shown as blue, orange, dark red and pink points, respectively. The formal 1σ (standard deviation) photometric uncertainties on the Pan-STARRS, WISE, and Spitzer points are smaller than the symbol size. Four different SED models are shown as solid curves: a pure hydrogen atmosphere model (red), a 50% hydrogen, 50% helium model (blue), a pure helium model (gold), and a blackbody curve (black). None of the SED models capture all of the SED's features, but all four yield mostly consistent effective temperatures and stellar parameters.



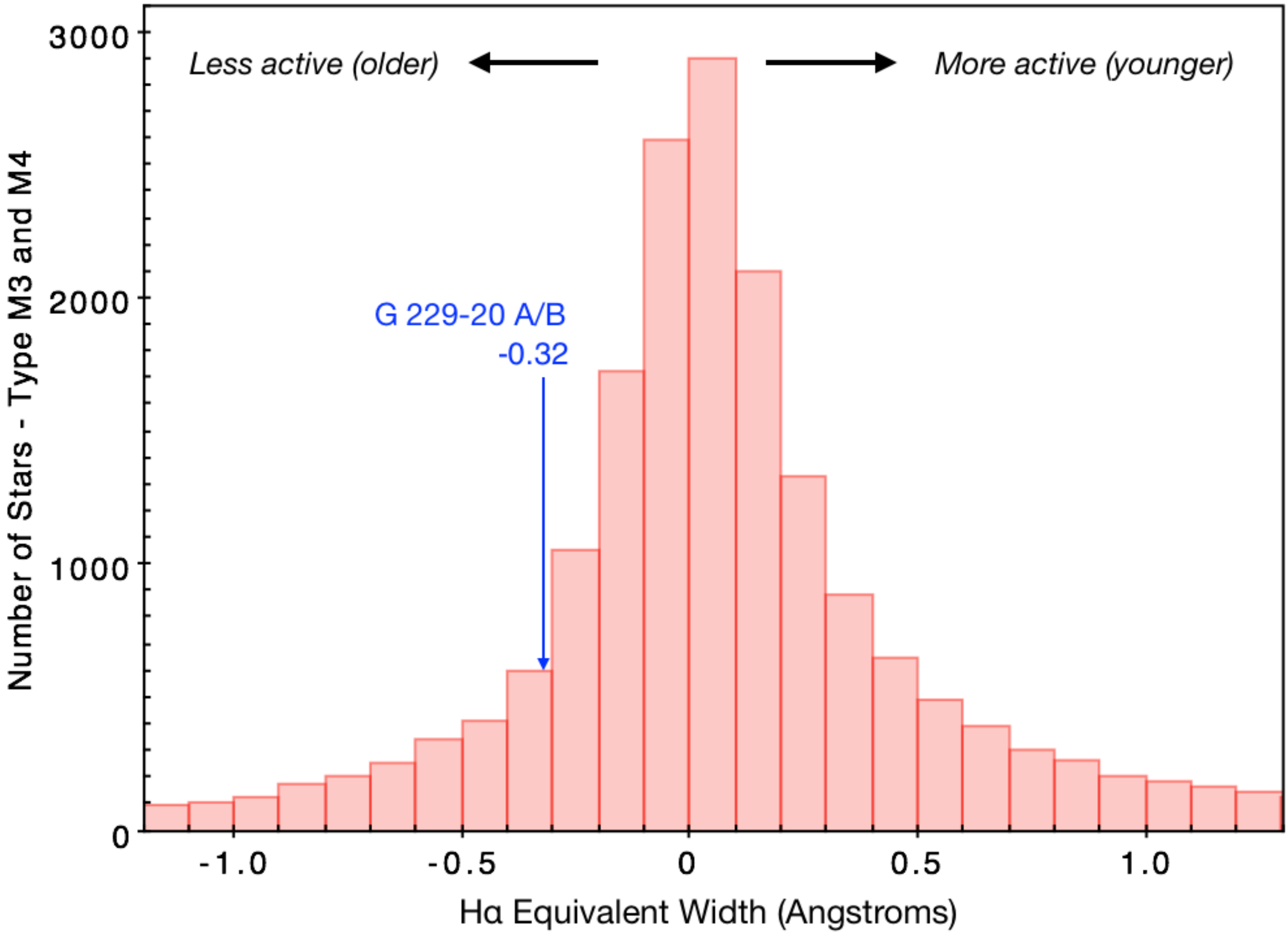
Extended Data Fig. 4 | Spectrum of WD 1856 near the Hα line. Our summed Hobby–Eberly/LRS2 spectrum (black connected points) is shown in comparison with three atmosphere models: a pure hydrogen model (red), a 50% hydrogen, 50% helium model (blue), and a pure helium model (gold). We disfavour a pure hydrogen atmosphere on the basis of our non-detection of an Hα feature in our LRS2 spectra, but otherwise remain uncertain about the precise composition of the envelope of WD 1856.



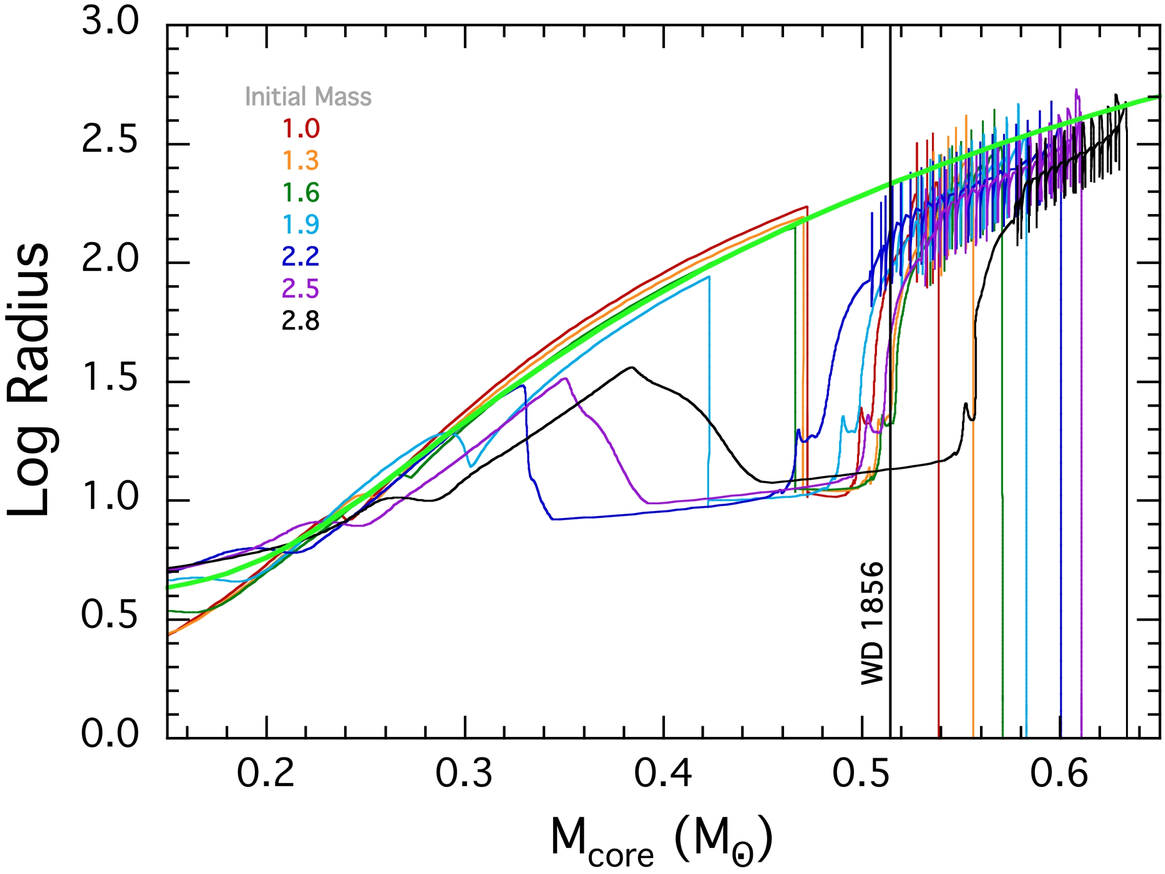
Extended Data Fig. 5 | Posterior probability distributions of transit parameters. This 'corner-plot' shows correlations between pairs of parameters in our MCMC transit fit (with circular orbits enforced) and histograms of the marginalized posterior probability distributions for each parameter. For clarity, we have plotted correlations with the inclination angle *i* instead of the fit parameter cos*i* and subtract the median time of transit (t_t). The orbital inclination *i*, scaled semimajor axis a/R_*, and the planet–star radius ratio R_p/R_* are strongly correlated, owing to the grazing transit geometry, but constrained by the prior on the stellar density. We do not include rows for the GTC and Spitzer photometric jitter terms because these are nuisance parameters that showed no correlation with the other physical parameters.



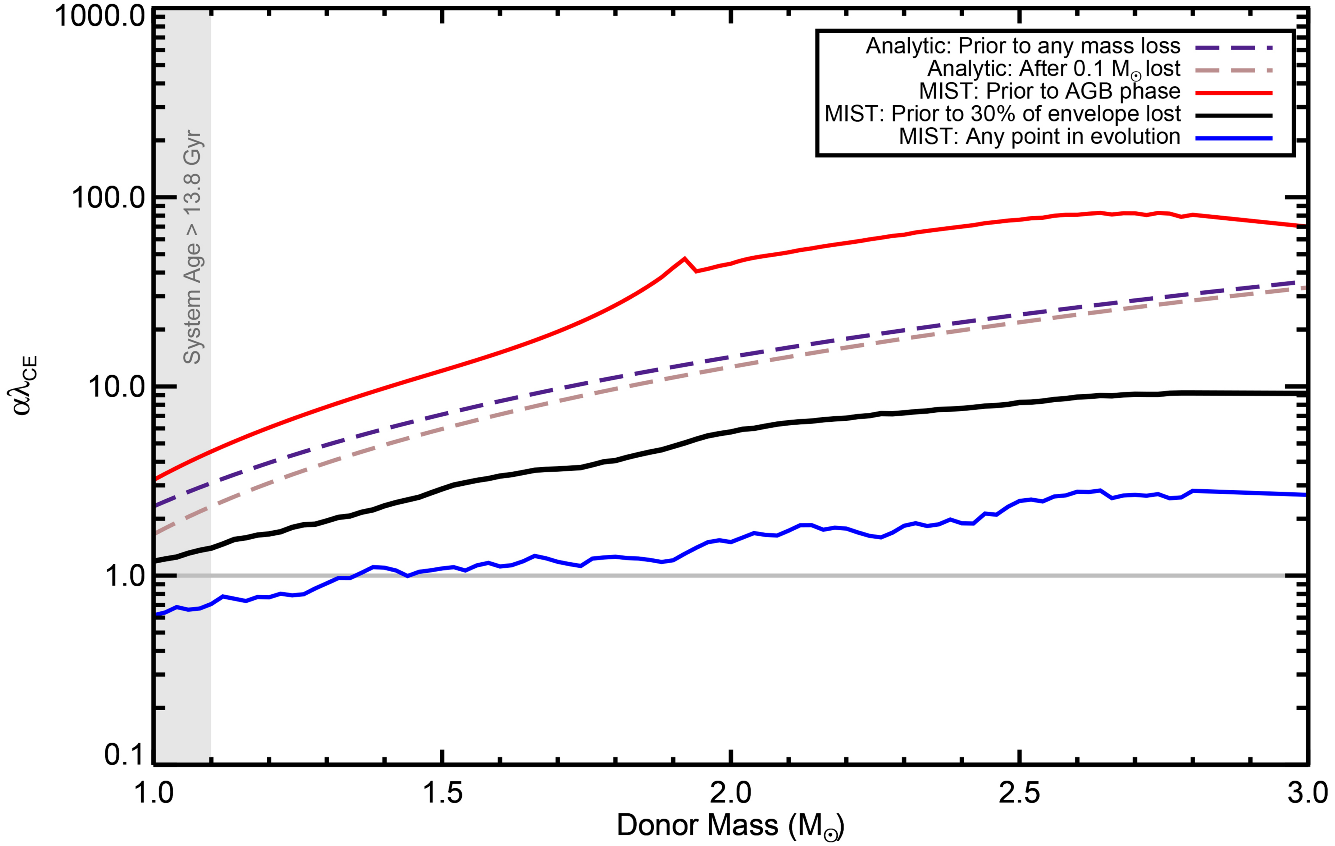
Extended Data Fig. 6 | Posterior probability distributions of transit parameters when eccentric orbits are allowed. This 'corner-plot' shows correlations between pairs of parameters in our MCMC transit fit (allowing eccentric orbits) and histograms of the marginalized posterior probability distributions for each parameter. This plot shows a subset of the parameters that correlate with the orbital eccentricity. For clarity, we have plotted correlations with the eccentricity e, argument of periastron ω and orbital inclination i instead of the fit parameters $\sqrt{e}\cos\omega$, $\sqrt{e}\sin\omega$ and δ.



Extended Data Fig. 7 | Hα equivalent width for G 229-20 A/B compared to other nearby M dwarfs. The histogram shows the Hα equivalent widths for a large sample of M dwarfs with similar spectral types from the Sloan Digital Sky Survey[103]. G 229-20 A/B (shown as a blue arrow) has a lower than average Hα equivalent width, but falls well within the distribution of field M dwarfs.



Extended Data Fig. 8 | Theoretical relationships between the star's radius and the mass of its core. We show MIST[120] evolution tracks in the radius–core mass plane for solar composition models with masses ranging from $1M_\odot$–$2.8M_\odot$. The RGB phase is clearly identifiable for core masses between $0.2M_\odot$ and $0.47M_\odot$, whereas the thermal pulses on the AGB are readily recognized at higher core masses of $\geq 0.5M_\odot$. The lime-green curve is the analytic expression given by equation (8). The vertical lines for each star mark the point where the envelope has been exhausted by the AGB wind.



Extended Data Fig. 9 | The minimum value of the efficiency parameter $\alpha\lambda_{CE}$ required for WD 1856 b to form via common-envelope evolution as a function of the progenitor stellar mass. The two dashed curves show the minimum $\alpha\lambda_{CE}$ values from our analytic calculation (equation (11)) required for a $15M_J$ object to eject the primary star's envelope. The purple dashed curve is taken directly from equation (11), and the brown dashed curve results if the progenitor star has lost $0.1M_{\odot}$ in a stellar wind by the time of the common envelope. The three solid curves show the minimum $\alpha\lambda_{CE}$ computed directly from MIST tracks in three different situations: before the star reaches the AGB (red), before more than 30% of the star's envelope mass has been lost (black), and at any point in the star's evolution, regardless of the mass lost (blue). Stars in the grey region at low masses evolve too slowly for the system to have left the main sequence more than 5.85 Gyr ago and are not viable solutions. For values of $\alpha\lambda_{CE} > 1$ (horizontal grey line), one must invoke the internal energy of the star to help to unbind the envelope during the common-envelope phase. Before mass is lost during the AGB phase, it is difficult for WD 1856 b to eject the common envelope, but it is possible that WD 1856 b could have ejected its progenitor's envelope if the common-envelope phase began after the progenitor reached the AGB. We have smoothed the lower two curves to remove some unphysical scatter that is probably due to numerical artefacts in the model grids.

Extended Data Table 1 | Comparison of white dwarf parameters from different atmosphere models

Parameter	**100% H**	**100% He**	**50%/50% H/He**
Mass, M_*	$(0.537 \pm 0.018)M_\odot$	$(0.396 \pm 0.018)M_\odot$	$(0.518 \pm 0.018)M_\odot$
Radius, R_*	$(0.0131 \pm 0.0014)R_\odot$	$(0.01489 \pm 0.0003)R_\odot$	$(0.0131 \pm 0.0003)R_\odot$
Surface gravity, $\log(g_{\mathrm{cgs}})$	7.931 ± 0.030	7.686 ± 0.030	7.915 ± 0.030
Effective temperature, T_{eff}	4,785 ± 60 K	4,430 ± 60 K	4,710 ± 60 K
Cooling age, t_{cool}	5.7 ± 0.5 Gyr	4.25 ± 0.5 Gyr	5.85 ± 0.5 Gyr

The superconducting quasicharge qubit

https://doi.org/10.1038/s41586-020-2687-9

Received: 13 December 2019

Accepted: 8 July 2020

Ivan V. Pechenezhskiy[1,3], Raymond A. Mencia[1,3], Long B. Nguyen[1], Yen-Hsiang Lin[1,2] & Vladimir E. Manucharyan[1]✉

The non-dissipative nonlinearity of Josephson junctions[1] converts macroscopic superconducting circuits into artificial atoms[2], enabling some of the best-controlled qubits today[3,4]. Three fundamental types of superconducting qubit are known[5], each reflecting a distinct behaviour of quantum fluctuations in a Cooper pair condensate: single-charge tunnelling (charge qubit[6,7]), single-flux tunnelling (flux qubit[8]) and phase oscillations (phase qubit[9] or transmon[10]). Yet, the dual nature of charge and flux suggests that circuit atoms must come in pairs. Here we introduce the missing superconducting qubit, 'blochnium', which exploits a coherent insulating response of a single Josephson junction that emerges from the extension of phase fluctuations beyond 2π (refs. [11–14]). Evidence for such an effect has been found in out-of-equilibrium direct-current transport through junctions connected to high-impedance leads[15–19], although a full consensus on the existence of extended phase fluctuations is so far absent[20–22]. We shunt a weak junction with an extremely high inductance—the key technological innovation in our experiment—and measure the radiofrequency excitation spectrum as a function of external magnetic flux through the resulting loop. The insulating character of the junction is manifested by the vanishing flux sensitivity of the qubit transition between the ground state and the first excited state, which recovers rapidly for transitions to higher-energy states. The spectrum agrees with a duality mapping of blochnium onto a transmon, which replaces the external flux by the offset charge and introduces a new collective quasicharge variable instead of the superconducting phase[23,24]. Our findings may motivate the exploration of macroscopic quantum dynamics in ultrahigh-impedance circuits, with potential applications in quantum computing and metrology.

Is a Josephson tunnel junction between two superconductors a superconducting link or an insulating break? Josephson showed[1] that a junction can be viewed as a nonlinear inductance that carries flux of $(\hbar/2e)\varphi$ and energy $E = -E_J\cos\varphi$, where φ is the superconducting phase difference, $\hbar$ is the reduced Plank constant, $2e$ is the charge of the Cooper pair and E_J is the Josephson energy. If quantum fluctuations of φ are small compared to 2π, the inductance can be linearized and the junction responds as a superconductor (Fig. 1a). Yet, an opposite scenario was suggested for the case in which φ is free to extend beyond the 2π interval[13]. In the following analysis, it is essential to take into account the intrinsic oxide capacitance of the junction across the Josephson element (Fig. 1b). The resulting circuit equations mimic an electron in a crystal (Fig. 1b): the flux is the position, the capacitance is the mass, the charge on the capacitor is the momentum and the Josephson energy corresponds to a periodic crystal field (Fig. 1c). The dynamics of φ can be described by extended Bloch waves and continuous Bloch bands. The energy $E_B(q)$ within the lowest band would be a $2e$-periodic function of the circuit analogue of quasimomentum, the quasicharge q. The quasicharge is the externally supplied charge and Cooper pair tunnelling is analogous to Bragg reflection. In other words, at low frequencies the junction transforms into a nonlinear Bloch capacitance (an equivalent of the effective mass) that stores charge q and is characterized by a $2e$-periodic charging energy $E_B(q)$. When quantum fluctuations of q are suppressed, the Bloch capacitance can be linearized, and hence the junction responds like an insulator.

The external circuit of the junction has a decisive role in choosing between two antagonistic scenarios. Bloch oscillations at frequency $I/2e$ are expected in response to a d.c. current $I = \dot{q}$ driven through the Bloch capacitance by an infinite-impedance current source[13]. By contrast, Josephson oscillations at a frequency of $2eV/\hbar$ are induced in response to a d.c. voltage $V = (\hbar/2e)\dot{\varphi}$ across the junction, applied with a zero-impedance voltage source. We short-circuit a Bloch capacitance with a large-value linear inductance. The resulting nonlinear and non-dissipative electrical circuit is the artificial atom blochnium (Fig. 1d, top). Blochnium is the dual of the transmon, a Josephson inductance shunted by a large-value linear capacitance (Fig. 1d, bottom). The high- (low-) impedance linear circuit element in blochnium (transmon) suppresses quantum fluctuations of the q– (φ–) variable and thereby stabilizes the insulating (superconducting) behaviour of the junction. Quasicharge localization profoundly differs from the usual Coulomb blockade: quasicharge cannot be offset from the mean value $q = 0$ by a static electric field thanks to complete screening by the galvanic shunt. The low-energy excitations of blochnium are anharmonic vibrations of quasicharge through the small junction, the spectrum of which is the focus of our work.

---

[1]Department of Physics, University of Maryland, College Park, MD, USA. [2]Present address: National Tsinghua University, Hsinchu, Taiwan. [3]These authors contributed equally: Ivan V. Pechenezhskiy, Raymond A. Mencia. ✉e-mail: vmanuchar@gmail.com

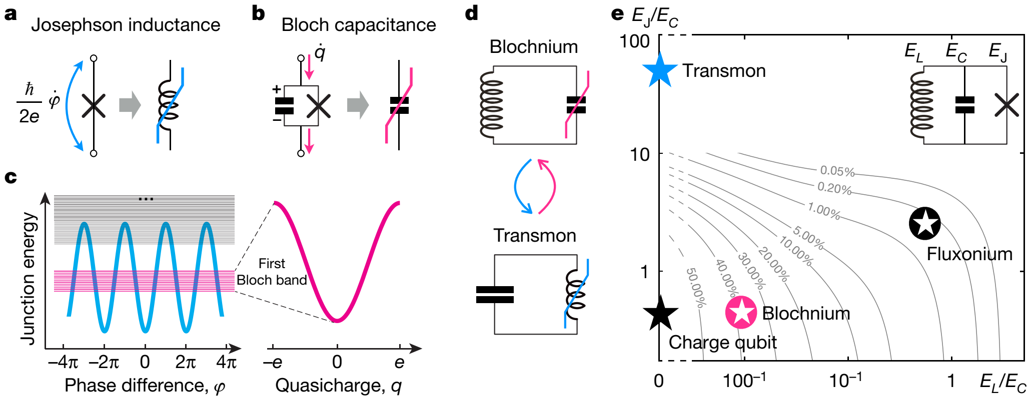


Fig. 1 | Blochnium artificial atom. a, A Josephson junction is a nonlinear inductance storing flux of $(\hbar/2e)\varphi$. **b**, A junction shunted by a small linear capacitance becomes a nonlinear Bloch capacitance storing quasicharge q. **c**, Spectrum of Bloch bands (magenta, grey) originating from the quantum motion of φ in the periodic Josephson potential (blue). The first Bloch band energy is a $2e$-periodic function of quasicharge q (magenta) and it defines the charging energy of the Bloch capacitance. **d**, Blochnium circuit concept (top) and its dual transmon circuit concept (bottom). **e**, Parameter space of the four fundamental qubits, which are all defined by the same three-element circuit (inset) with vastly different combinations of E_L, E_C and E_J (see text). The contours show the calculated probability of $|\varphi| > \pi$.

Constructing blochnium amounts to choosing three elementary energy scales: the Josephson energy E_J, the charging energy $E_C = e^2/(2C)$ of the total capacitance C across the Josephson element and the inductive energy $E_L = (\hbar/2e)^2/L$ associated with storing a flux quantum in the inductance L (Fig. 1e). Our devices have $E_J/E_C \approx 1$ and $E_L/E_C \approx 1/100$. The first condition conveniently maximizes the width of the lowest Bloch band, along with the gap to the next one. As for the second condition, in the special case of $E_L = 0$, we are left with two isolated grains linked by a single Cooper pair tunnelling, that is, a charge qubit. In such a case, the quasicharge loses dynamics and it can be interpreted as an external charge offset. For $E_L/E_C \lesssim 1$, we get the fluxonium[25,26], a high-coherence implementation of a flux qubit, in which a superconducting loop is disrupted by the tunnelling of a single flux quantum. In a charge qubit, reducing E_C proliferates multiple Cooper pair tunnelling[10], which establishes a well defined phase difference across the junction owing to the Heisenberg uncertainty principle. This is how a charge qubit evolves into a transmon for $E_J/E_C \approx 10$–100. Blochnium emerges from the dual evolution of the fluxonium on reducing E_L. The probability to find φ outside a single Josephson well becomes about 1 for $E_L/E_C \lesssim 1/100$ (Fig. 1e).

One of our measured devices is shown in Fig. 2a. Building on fluxonium results, we constructed a compact shunt using the kinetic inductance of a Josephson junction chain (inset of Fig. 2a, blue). The key innovation here is that we released the entire circuit from the substrate and suspended it in vacuum (Fig. 2a). With optimally chosen junction parameters, Josephson chains can reach an exceptionally high inductance density of $10^4\mu_0$ (~10 mH m^{-1}), where μ_0 is the vacuum permeability, before the onset of the detrimental effects associated with the superconductor–insulator transition[27,28]. However, the total inductance is also limited by the chain self-capacitance, originating from the stray electrostatic coupling between the opposite-facing metal islands (inset of Fig. 2a, red). Besides introducing parasitic modes, the self-capacitance contributes to C, and this effect prevents reducing E_L/E_C to values much lower than unity by lengthening the chain. The stray capacitance, however, is unnecessarily large in most superconducting circuits owing to the high relative dielectric permittivity of silicon ($\varepsilon \approx 12$) and sapphire ($\varepsilon \approx 10$)—the typical low-loss substrate materials. By eliminating the substrate contribution, we reduced the stray capacitance nearly ten-fold, which provided the required leap in the inductance value.

The substrate-free blochnium devices in Fig. 2 are created in a two-step fabrication process. First, a superconducting loop with up to 460 Al/AlO$_x$/Al chain junctions and one small junction is fabricated using the standard Dolan bridge technique. Next, a gentle burst of isotropic silicon etch is applied, with the oxidized Al film acting as a natural mask. Because silicon etching is more efficient underneath the skinnier leads, the small-junction end of the chain (labelled ‘1’ in Fig. 2a) detaches from the substrate before the other parts and immediately curls upwards driven by the strain relaxation. The curling effect is robust and reproducible, and it is possible to vary the amount of curling (Fig. 2b, c). We focused on devices with a nearly vertically standing chain (Fig. 2a), in which parasitic capacitance is minimal.

The loop is inductively coupled to the readout circuitry following a previously developed method[29]. A small section of the loop (labelled ‘2’ in Fig. 2a) is connected to a capacitive antenna (with leads that are labelled ‘3’ and ‘4’ in Fig. 2a), which forms a readout resonator for coupling the device to the measurement apparatus. The transition spectrum as a function of the flux through the loop (Fig. 3a) was measured using conventional two-tone radiofrequency spectroscopy[30] (see Methods for spectroscopy details). To identify transitions, we compared the data (Fig. 3a, markers) to the spectrum of a three-element circuit (Fig. 1e) Hamiltonian

$$H_1 = 4E_C\left(\frac{Q}{2e}\right)^2 + \frac{1}{2}E_L\varphi^2 - E_J\cos(\varphi - \varphi_{\mathrm{ext}}). \tag{1}$$

The Hamiltonian of equation (1) describes a particle that is both in a flux-tuned periodic potential and a soft harmonic trap due to the E_L term. The operators φ and Q obey the position–momentum-type commutation relation $[\varphi, Q/(2e)] = \mathrm{i}$. The simple model of equation (1) accurately fits the lowest five transitions (Fig. 3, dashed lines). The fit parameters $E_J/h = 4.70$ GHz, $E_C/h = 7.07$ GHz and $E_L/h = 66.5$ MHz indeed define a previously inaccessible spot on the circuit parameter map of Fig. 1e ($E_J/E_C = 0.66$, $E_L/E_C = 0.009$). The capacitance $C \approx 2.7$ fF can be almost entirely accounted for by the oxide capacitance of the small junction. The inductance $L \approx 2.5$ μH exceeds that of a typical fluxonium ten-fold but shows no influence of parasitic modes within the entire frequency range of Fig. 3.

The rapid crossover in the flux-modulation characteristic of the transitions from a weak harmonic to a strong saw-tooth (Fig. 3) has no analogues among previously reported spectra of superconducting quantum interference devices. We introduce a phenomenological model of an inductively shunted Bloch capacitance (Fig. 1d) with Hamiltonian

$$H_2 = 2\pi^2 E_L[m - (\varphi_{\mathrm{ext}}/2\pi)]^2 + E_B(q). \tag{2}$$

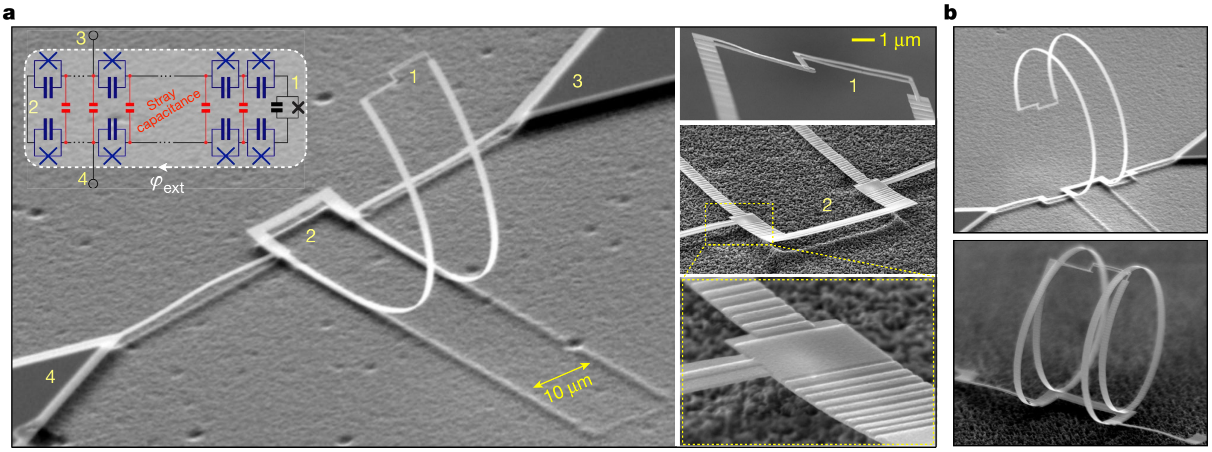


Fig. 2 | Device implementation. a, Scanning electron micrographs of a fabricated device released from the substrate to reduce the stray capacitance. The released Josephson chain curls upwards and elevates the small junction by a few tens of micrometres above the substrate. The inset shows a circuit model of the device in the form of a superconducting loop interrupted by a small-area junction (black) and the larger-area chain junctions (dark blue). The stray capacitances are marked in red. The indexes 1–4 mark the small junction, the opposite end of the loop and the two connections to the readout circuitry (not shown in the image), respectively. An external magnetic field induces flux φ_{ext} through the loop. **b**, Examples of released blochnium circuits with different degrees of curling. The two parts of the Josephson junction chain are spaced by 10 µm in all devices.

Here the quasicharge q transferred across the shunt is a compact variable at the interval $(-e, e]$ and $E_B(q)$ is the $2e$-periodic charging energy of the Bloch capacitance. The conjugate momentum m is an integer operator satisfying $e^{-i\pi q/e} m e^{i\pi q/e} = m+1$. The external flux φ_{ext} couples to momentum like a gauge field. For a sufficiently large E_L, the momentum m counts the flux quanta (or the 2π slips of phase) in the loop, although such a notion becomes progressively more vague upon reducing E_L into the regime of interest.

Aside from the qualitatively unimportant effect of the higher harmonics of $E_B(q)$, the Hamiltonian of equation (2) models a quantum pendulum with a deflection angle of $\pi q/e$. The same pendulum model describes transmons, where the deflection angle is φ and the external flux is replaced by the offset charge, thereby providing a quantitative basis for the duality. Using the circuit parameters E_J and E_C, extracted above by fitting the H_1 model to the data, we calculate dispersion of the Bloch bands originating from the small junction in our circuit. The function $E_B(q)$ in H_2 is set to be the lowest band, and the higher bands are disregarded. Using the H_1 fit value for E_L, we numerically diagonalize the H_2 and use both models to interpret the spectroscopy results (Fig. 4).

The lowest five energy states exhibit all the essential features of a transmon device[10] if one considers $\varphi_{ext}/(2\pi)$ to be the offset charge normalized by $2e$ (Fig. 4a). The ground state $|0\rangle$ has a vanishing flux dispersion (unresolvable in Fig. 4a), corresponding to a hardly measurable persistent current of 7 pA (Methods and Extended Data Fig. 2). The $|0\rangle$ energy level lies deep inside the first Bloch band (Fig. 4b), and this property links the absence of magnetic-field response to the localization of quasicharge around $q = 0$. Indeed, with $\langle 0|[q/(2e)]^2|0\rangle \approx 0.019$, the quasicharge wavefunction has exponentially small tails at the Brillouin zone edges $|q| = e$ (Fig. 4c), in which case φ_{ext} can be eliminated from equation (2) by a gauge transformation of $\exp[(i\varphi_{ext} q/(2e)]$. The $|1\rangle$ level also lies well inside the first Bloch band, having only 5% of flux modulation due to a reduced quasicharge confinement. The $|0\rangle \rightarrow |1\rangle$ transition corresponds to semi-classical oscillations of quasicharge inside the Bloch band potential. The flux quantization recovers already for states $|3\rangle$ and $|4\rangle$ as the quasicharge spills over the entire Brillouin zone. At higher energies, the spectra of the Hamiltonians of equations (1) and (2) deviate owing to the presence of higher Bloch bands, which are ignored in the model of equation (2) (Fig. 4b). In fact, this quantitative discrepancy confirms that our Bloch capacitance emerges from the physics of the underlying Josephson effect.

Quasicharge localization with the root-mean-square value of 13% of a Cooper pair (Fig. 4c) justifies the Bloch band picture, in which the junction responds as a Bloch capacitance rather than a Josephson inductance. The phase difference φ across the junction extends beyond a single Josephson well but it remains localized at the scale of a few wells (Fig. 4d). Thus, φ is no longer compact at $(-\pi, \pi]$, and its localization length continually increases with L. Notably, the quantity $\langle 0|\cos\varphi|0\rangle$ remains non-zero in the limit $L \rightarrow \infty$. This means that while the junction responds as a capacitance, Cooper pairs can virtually tunnel back and

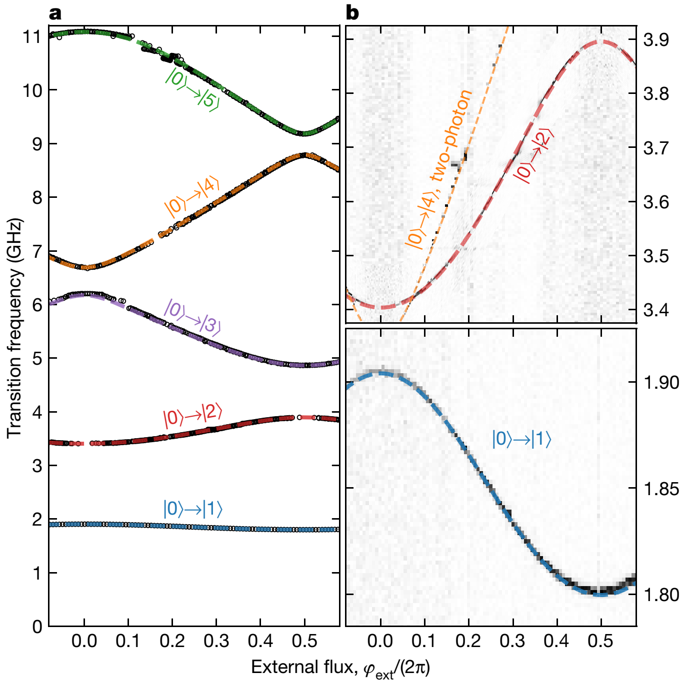


Fig. 3 | Measured transitions of blochnium. a, Transition frequencies (black markers) extracted from the two-tone spectroscopy data as a function of the external flux through the loop, and the fit (dashed lines) to the spectrum of the Hamiltonian of equation (1). **b**, Raw data zoom-in on the $|0\rangle \rightarrow |2\rangle$ and the two-photon $|0\rangle \rightarrow |4\rangle$ transitions (top) and the qubit transition $|0\rangle \rightarrow |1\rangle$ (bottom). The experimental error is much smaller than the marker size. Note that the modulation of the qubit flux is only about 100 MHz.

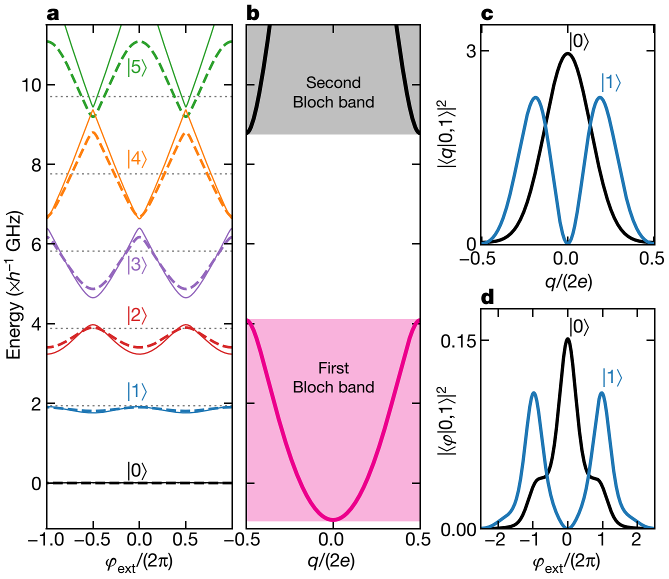


Fig. 4 | Interpretation of spectroscopy data. a, Eigenenergies of the Hamiltonians of equation (1) (dashed lines) and equation (2) (solid lines), calculated using the extracted device parameters, versus the external flux φ_{ext}. Dotted lines indicate the spectrum of equation (1) for $E_J = 0$. **b**, Calculated energies in the lowest two Bloch bands as a function of the quasicharge q. **c**, **d**, Probability distributions of the ground and first excited states at $\varphi_{ext} = 0$ in the quasicharge representation (**c**) and in the phase difference representation (**d**).

forth across the oxide. In fact, such processes are responsible for the nonlinearity of the Bloch capacitance $C_B(q) = (d^2E_B(q)/dq^2)^{-1}$. In addition, virtual Cooper pair tunnelling would increase $C_B(0)$ considerably above C in junctions with $E_J \gg E_C$. In our device, $C_B(0) \approx C$, which allows a straightforward illustration of the insulating character of the junction. Namely, setting $E_J = 0$ and keeping the fit values of E_C and E_L reproduces (with an accuracy of a few per cent) both the $|0\rangle \rightarrow |1\rangle$ transition frequency and the matrix element $\langle 0|\varphi|1\rangle$ (Methods and Extended Data Fig. 3). In other words, the low-energy dynamics of our device is consistent with simply removing the Josephson element from the circuit.

Achieving the fluctuations regime of the decompactified phase difference and the localized quasicharge allowed us to complete the table of fundamental superconducting artificial atoms with blochnium. The finite-L shunt eliminates the offset charge sensitivity of blochnium transitions, whereas the extension of phase fluctuations beyond the $(-\pi, \pi]$ interval renders the $|0\rangle \rightarrow |1\rangle$ qubit transition practically unaffected by the background level of flux noise[8]. Moreover, transitions to non-computational states are flux-tunable and anharmonic, which is a desirable resource for quantum engineering. Initial time-domain measurements of our substrate-free devices revealed an energy relaxation time of $T_1 \approx 10$ μs and a relaxation-limited spin-echo coherence time of $T_2 \approx 20$ μs (Methods and Extended Data Fig. 6).

The blochnium qubit is enabled by a remarkable circuit element to which we refer as hyperinductance: a lossless linear inductance of $L \approx 2.5$ μH operating beyond the frequency of $\omega/(2\pi) \approx 13$ GHz, such that $L\omega > 200$ kΩ. This impedance is a factor of 30 greater than the resistance quantum for Cooper pairs $h/(2e)^2$, and it is probably the highest characteristic impedance attained so far by an electromagnetic structure. Among other applications, hyperinductance has long been sought after for realizing fault-tolerant logical operations on superconducting qubits[31–33] and for implementing the quantum current standard via Bloch oscillations[34,35].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2687-9.

1. Josephson, B. D. Possible new effects in superconductive tunnelling. *Phys. Lett.* **1**, 251–253 (1962).
2. Clarke, J., Cleland, A., Devoret, M. H., Esteve, D. & Martinis, J. Quantum mechanics of a macroscopic variable: the phase difference of a Josephson junction. *Science* **239**, 992–997 (1988).
3. Devoret, M. H. & Schoelkopf, R. Superconducting circuits for quantum information: an outlook. *Science* **339**, 1169–1174 (2013).
4. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
5. Clarke, J. & Wilhelm, F. K. Superconducting quantum bits. *Nature* **453**, 1031–1042 (2008).
6. Nakamura, Y., Pashkin, Y. A. & Tsai, J. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
7. Vion, D. et al. Manipulating the quantum state of an electrical circuit. *Science* **296**, 886–889 (2002).
8. Chiorescu, I., Nakamura, Y., Harmans, C. J. P. M. & Mooij, J. E. Coherent quantum dynamics of a superconducting flux qubit. *Science* **299**, 1869–1871 (2003).
9. Martinis, J. M., Nam, S., Aumentado, J. & Urbina, C. Rabi oscillations in a large Josephson-junction qubit. *Phys. Rev. Lett.* **89**, 117901 (2002).
10. Koch, J. et al. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
11. Schmid, A. Diffusion and localization in a dissipative quantum system. *Phys. Rev. Lett.* **51**, 1506 (1983).
12. Bulgadaev, S. A. Phase diagram of a dissipative quantum system. *JETP Lett.* **39**, 264–267 (1984).
13. Averin, D. V., Zorin, A. B. & Likharev, K. K. Bloch oscillations in small Josephson junctions. *Sov. Phys. JETP* **61**, 407–413 (1985).
14. Schön, G. & Zaikin, A. D. Quantum coherent effects, phase transitions, and the dissipative dynamics of ultra small tunnel junctions. *Phys. Rep.* **198**, 237–412 (1990).
15. Kuzmin, L. S. & Haviland, D. B. Observation of the Bloch oscillations in an ultrasmall Josephson junction. *Phys. Rev. Lett.* **67**, 2890 (1991).
16. Haviland, D. B. & Delsing, P. Cooper-pair charge solitons: the electrodynamics of localized charge in a superconductor. *Phys. Rev. B* **54**, R6857–R6860 (1996).
17. Penttilä, J. S., Parts, Ü., Hakonen, P. J., Paalanen, M. A. & Sonin, E. B. "Superconductor–insulator transition" in a single Josephson junction. *Phys. Rev. Lett.* **82**, 1004 (1999).
18. Watanabe, M. & Haviland, D. B. Coulomb blockade and coherent single-Cooper-pair tunneling in single Josephson junctions. *Phys. Rev. Lett.* **86**, 5120 (2001).
19. Corlevi, S., Guichard, W., Hekking, F. W. J. & Haviland, D. B. Phase-charge duality of a Josephson junction in a fluctuating electromagnetic environment. *Phys. Rev. Lett.* **97**, 096802 (2006).
20. Ergül, A. et al. Localizing quantum phase slips in one-dimensional Josephson junction chains. *New J. Phys.* **15**, 095014 (2013).
21. Cedergren, K. et al. Insulating Josephson junction chains as pinned luttinger liquids. *Phys. Rev. Lett.* **119**, 167701 (2017).
22. Murani, A. et al. Absence of a dissipative quantum phase transition in Josephson junctions. *Phys. Rev. X* **10**, 021003 (2020).
23. Matveev, K. A., Larkin, A. I. & Glazman, L. I. Persistent current in superconducting nanorings. *Phys. Rev. Lett.* **89**, 096802 (2002).
24. Koch, J., Manucharyan, V., Devoret, M. H. & Glazman, L. I. Charging effects in the inductively shunted Josephson junction. *Phys. Rev. Lett.* **103**, 217004 (2009).
25. Manucharyan, V. E., Koch, J., Glazman, L. I. & Devoret, M. H. Fluxonium: single Cooper-pair circuit free of charge offsets. *Science* **326**, 113–116 (2009).
26. Nguyen, L. B. et al. High-coherence fluxonium qubit. *Phys. Rev. X* **9**, 041041 (2019).
27. Manucharyan, V. E. et al. Evidence for coherent quantum phase slips across a Josephson junction array. *Phys. Rev. B* **85**, 024521 (2012).
28. Kuzmin, R. et al. Quantum electrodynamics of a superconductor-insulator phase transition. *Nat. Phys.* (2019).
29. Kou, A. et al. Simultaneous monitoring of fluxonium qubits in a waveguide. *Phys. Rev. Appl.* **9**, 064022 (2018).
30. Schuster, D. I. et al. ac Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. *Phys. Rev. Lett.* **94**, 123602 (2005).
31. Bell, M. T., Sadovskyy, I. A., Ioffe, L. B., Kitaev, A. Y. & Gershenson, M. E. Quantum superinductor with tunable nonlinearity. *Phys. Rev. Lett.* **109**, 137003 (2012).
32. Douçot, B. & Ioffe, L. B. Physical implementation of protected qubits. *Rep. Prog. Phys.* **75**, 072001 (2012).
33. Brooks, P., Kitaev, A. & Preskill, J. Protected gates for superconducting qubits. *Phys. Rev. A* **87**, 052306 (2013).
34. Nguyen, F. et al. Current to frequency conversion in a Josephson circuit. *Phys. Rev. Lett.* **99**, 187005 (2007).
35. Di Marco, A., Hekking, F. W. J. & Rastelli, G. Quantum phase-slip junction under microwave irradiation. *Phys. Rev. B* **91**, 184512 (2015).

Methods

Device fabrication

Devices were fabricated using the conventional Dolan bridge technique[36,37]. A bilayer of methylmethacrylate/polymethylmethacrylate was spun on top of a high-resistivity ($\rho > 10$ kΩ cm) silicon substrate covered by the native oxide before patterning the device with electron-beam lithography. The patterned device was then loaded into a electron-beam evaporator, where the Al/AlO$_x$/Al Josephson junctions were created by double-angle shadow evaporation of aluminium with an intermediate static oxidation step. The first layer of aluminium was 20 nm thick and the second one was 40 nm thick. After deposition, the resist bilayer was lifted off in a 60 °C acetone bath.

Once each device was created in a two-dimensional fashion, we performed an etch to release the aluminium circuit from the silicon substrate. The device was etched using a xenon difluoride reactive ion etching technique, which relies on the dry etchant selectivity of silicon over aluminium[38,39]. Each sample was etched for 2 min at a pressure of 2 mtorr. No noticeable detrimental effects were observed in the scanning electron micrographs of our etched devices after using this fabrication technique.

Measurements

The device readout was performed using an on-chip resonator that was inductively coupled[27] to the device loop via 10–18 Josephson junctions (these junctions, shared by the qubit loop and the on-chip resonator, are labelled ‘2’ in Fig. 2a). The resonator capacitance was provided by the bowtie-shaped antenna pads (partially visible in Fig. 2a, b). To improve the linearity of the readout resonator, the coupler junctions were designed to be twice as wide as those of the inductance chain (lower right inset in Fig. 2a). By placing the resonator antenna at the opposite side of the chain from the small Josephson junction, we minimized the contribution from the antenna pads to the capacitance of the small Josephson junction, increasing the device charging energy E_C.

For the spectroscopy measurements, each device was placed in a three-dimensional copper box with a broad-band microwave launcher at its single drive/readout port, effectively making the copper box a three-dimensional waveguide in the frequency range of the on-chip resonators. A home-made superconducting coil was placed externally to the copper waveguide to control the flux bias through the device loop. The spectroscopy data were obtained using conventional two-tone reflectometry in a dilution refrigerator at 13 mK, with the microwave setup closely resembling the one used in ref. [26].

Qubit spectroscopy

The two-tone spectroscopy data discussed in the main text are provided in Extended Data Fig. 1, where the fit to the transition spectrum of the Hamiltonian of equation (1) is superimposed on the data. The deviation between the data and the fit becomes noticeable only from the $|0\rangle \rightarrow |6\rangle$ transition around 12 GHz. However, when compared to the transition frequencies, these deviations remain small within the entire measurement frequency range. Upon close inspection, several anticrossings with two-level-like systems could be identified. The avoided level splittings, however, are small compared to those expected from splittings with parasitic modes.

Junction insulating character

The persistent current in the ground state deduced from the device parameters is shown in Extended Data Fig. 2. Remarkably, the maximum ground-state supercurrent is three orders of magnitude smaller than the critical current of the bare junction.

The flux dispersion of the lowest three energy states is shown in Extended Data Fig. 3a. Extended Data Fig. 3b shows the matrix elements $\langle 0|\varphi|1\rangle$ computed for the actual device and for the hypothetical device, in which the Josephson junction is eliminated. The values of the two matrix elements are close to each other and their difference does not exceed 4% across the entire flux quantum.

Extended Data Figs. 4, 5 further illustrate the ground-state properties by showing the ground-state wavefunction in different bases. In addition, Extended Data Fig. 4 demonstrates the expected evolution of the wavefunction upon increasing the shunting inductance, as does Extended Data Fig. 5 for a different ratio E_J/E_C.

Energy relaxation and decoherence

The energy relaxation time T_1 and coherence time T_2 were conventionally measured by observing the qubit population decay and by applying a spin-echo measurement sequence, respectively. An example of the measured time traces are shown in Extended Data Fig. 6. We observe that T_2 is comparable to $2T_1$, which suggests that the coherence times are currently limited by the energy relaxation processes. The detailed analysis of the energy relaxation and decoherence warrants a separate study that is beyond the scope of this work. However, we note that dephasing due to the first-order sensitivity of the qubit transition frequency to the flux noise should allow coherence times exceeding 300 μs across the entire flux quantum (Extended Data Fig. 7).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The codes that support the findings of this study are available from corresponding author upon reasonable request.

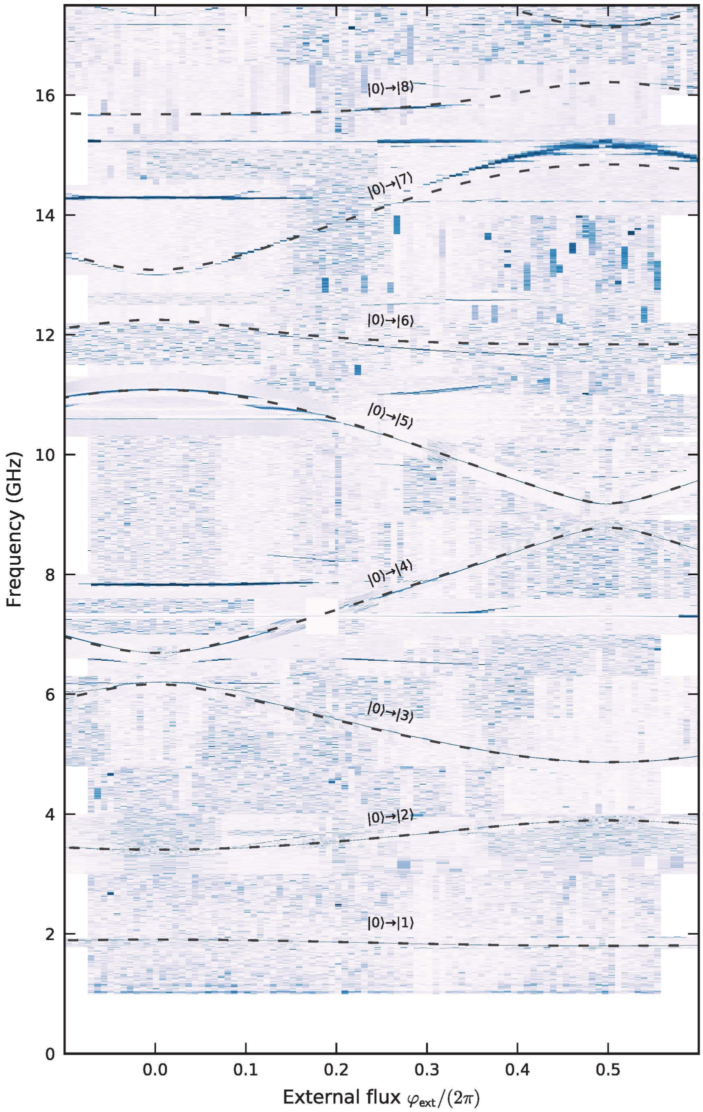
36. Dolan, G. J. Offset masks for lift-off photoprocessing. *Appl. Phys. Lett.* **31**, 337–339 (1977).
37. Frunzio, L., Wallraff, A., Schuster, D., Majer, J. & Schoelkopf, R. Fabrication and characterization of superconducting circuit QED devices for quantum computation. *IEEE Trans. Appl. Supercond.* **15**, 860–863 (2005).
38. Chang, F. I. et al. Gas-phase silicon micromachining with xenon difluoride. In *Proc. SPIE* 2641, https://doi.org/10.1117/12.220933 (SPIE, 1995).
39. Chu, Y. et al. Suspending superconducting qubits by silicon micromachining. *Appl. Phys. Lett.* **109**, 112601 (2016).

Acknowledgements We acknowledge funding from NSF-CAREER (1455261), Alfred P. Sloan Foundation, NSF PFC at JQI (1430094), and the ARO-LPS HiPS programme (W911NF-18-1-0146).

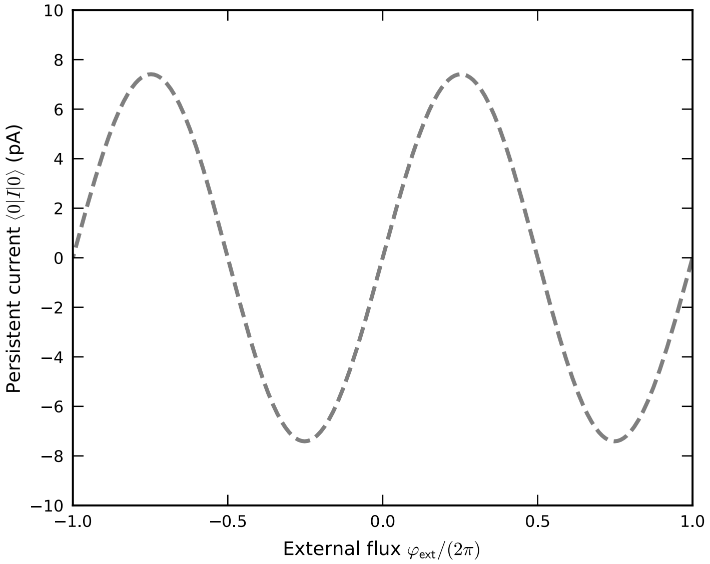
Author contributions R.A.M. fabricated devices and performed measurements guided by I.V.P.; I.V.P. analysed the data and co-wrote the manuscript with V.E.M.; L.B.N. and Y.-H.L. built the low-temperature microwave measurement setup; V.E.M. managed the project. All authors contributed to discussions of the results.

Competing interests The authors declare no competing interests.

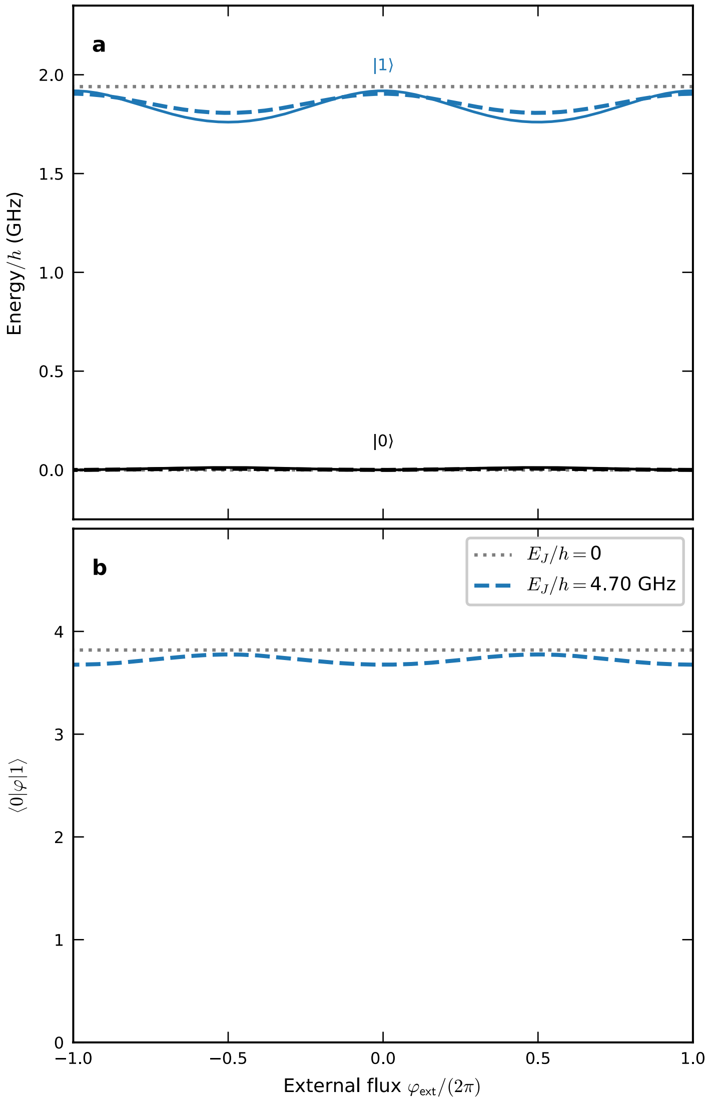
Additional information
Correspondence and requests for materials should be addressed to V.E.M.
Peer review information *Nature* thanks Göran Johansson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



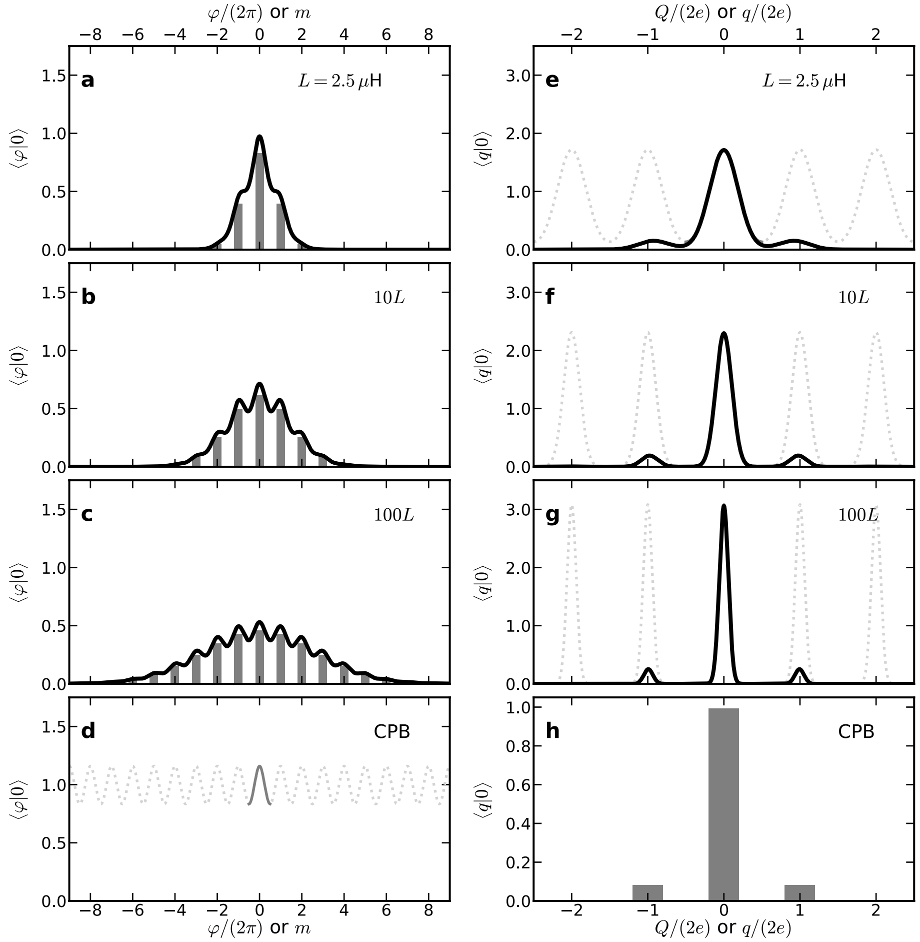
Extended Data Fig. 1 | Qubit spectroscopy. Stitched one- and two-tone spectroscopy data as a function of the spectroscopy frequency and the normalized external flux through the loop. The fit (dashed lines) to the transition spectrum of the Hamiltonian of equation (1) is superimposed on the data. The data were collected in a patch-wise manner, with the measurement parameters optimized locally to improve the visibility of the transitions out of the ground state $|0\rangle$. As in Fig. 3b, nonlinear colour maps are used to assign colour to the measured signal. Note that the deviation between the fit and the data is noticeable only from the $|0\rangle \rightarrow |6\rangle$ transition.



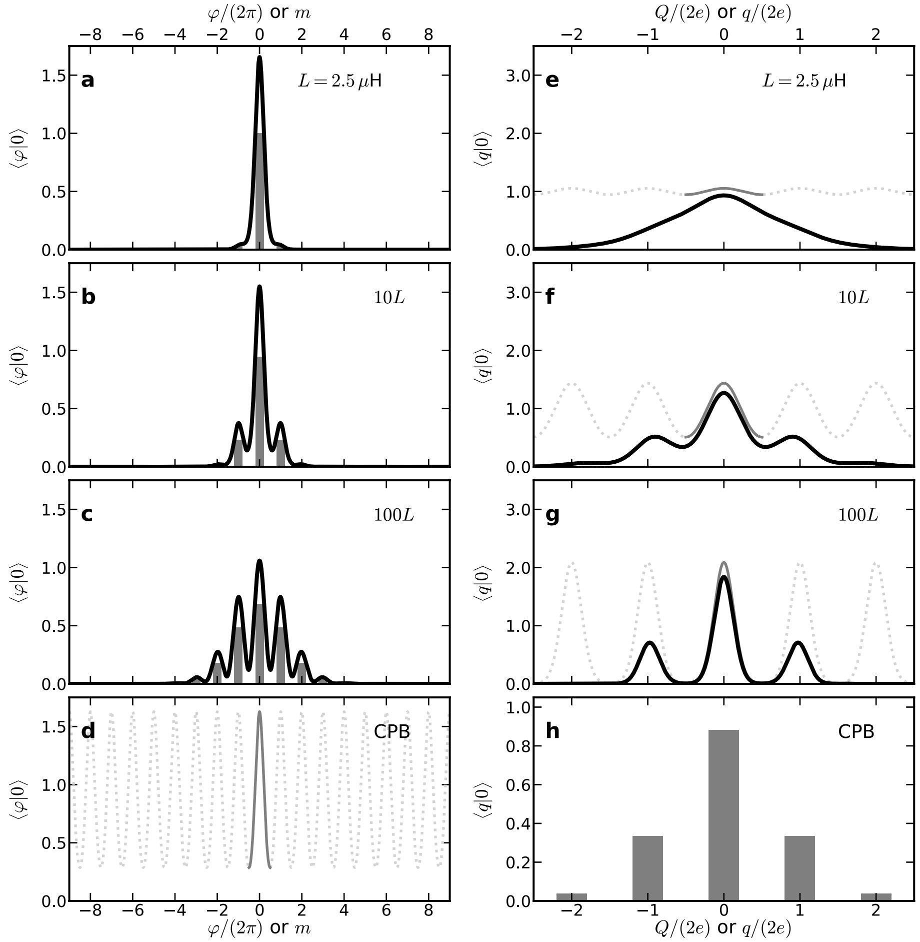
Extended Data Fig. 2 | Persistent current. Persistent current in the ground state of the device $\langle 0|I|0\rangle \equiv I_0\langle 0|\sin(\varphi - \varphi_{\text{ext}})|0\rangle$, where $I_0 = 9.5$ nA is the junction critical current, plotted as a function of the external flux, φ_{ext}. The current is calculated using the extracted device parameters.



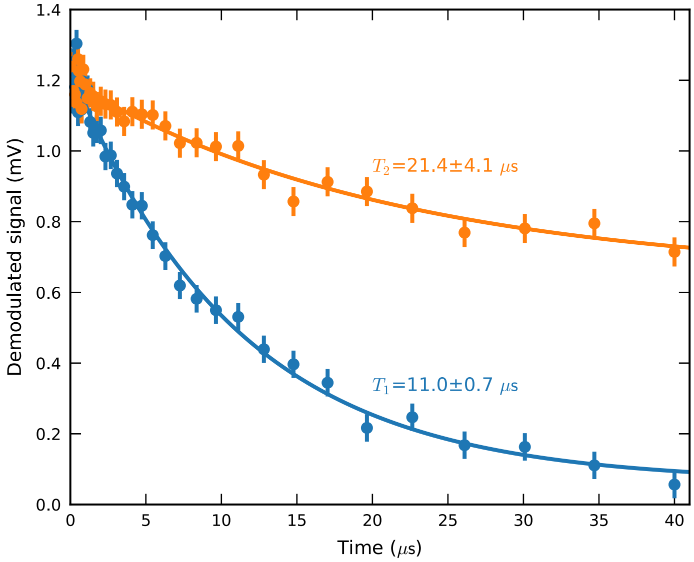
Extended Data Fig. 3 | Flux dispersion and matrix elements. a, Zoom-in on the lowest two states of Fig. 4a. Eigenenergies of the Hamiltonians of equation (1) (dashed lines) and equation (2) (solid lines) and of a hypothetical device without the Josephson junction (grey dotted lines) as a function of the external flux, $\varphi_{\rm ext}$. The spectra are calculated using the extracted device parameters. **b**, Matrix element $\langle 0|\varphi|1\rangle$ as a function of the external flux, $\varphi_{\rm ext}$. The dashed line corresponds to $E_J/h = 4.70$ GHz and should be compared to the dotted line, which corresponds to $E_J = 0$—that is, to the hypothetical case without the Josephson junction. In both panels, $E_C/h = 7.07$ GHz and $E_L/h = 66.5$ MHz.



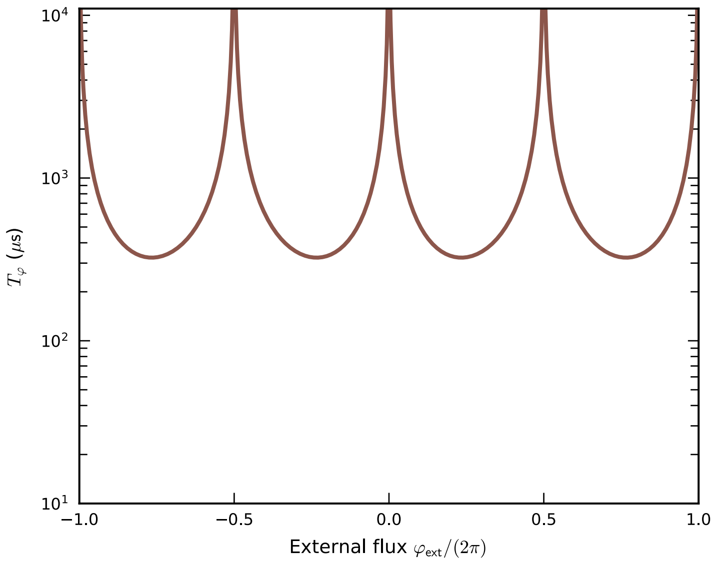
Extended Data Fig. 4 | Ground-state wavefunctions of the measured device. **a**–**h**, Ground-state wavefunctions in the phase (φ) and integer-flux (m) bases (**a**–**d**) and in the charge (Q) and quasicharge (q) bases (**f**–**h**). **a**, **e**, Ground-state wavefunctions of the device discussed in the main text for $\varphi_{\text{ext}} = 0$. The black solid line in **a** corresponds to the unbounded, continuous-phase φ basis and the black solid line in **e** to the continuous-charge Q basis, which are the natural bases for the Hamiltonian of equation (1). The stems in **a** correspond to the discrete integer-flux basis and the dotted grey line in **e** to the periodic quasicharge basis, which are the natural bases for the Hamiltonian of equation (2). $E_J/h = 4.70$ GHz, $E_C/h = 7.07$ GHz and $E_L/h = 66.5$ MHz. **b**, **f**, Same as **a**, **e**, but for a ten-times-larger inductance, that is, $E_L/h = 6.65$ MHz. **c**, **g**, Same as **a**, **e**, but for a 100- times-larger inductance, that is, $E_L/h = 0.67$ MHz. **d**, **h**, Cooper pair box (CPB) wavefunction in the phase φ (**d**) and charge Q (**h**) bases computed for the same values of E_J and E_C used in **a**–**c** and **e**–**g**, $E_J/E_C = 0.66$. A single period (−π, π] in **d** is highlighted in solid grey. The offset charge is set to zero.



Extended Data Fig. 5 | Ground-state wavefunctions for modified device parameters. Same as Extended Data Fig. 4, but for $E_J/h = 4.70$ GHz and $E_C/h = 1.18$ GHz, so that $E_J/E_C = 4.0$. Single periods $(-e, e]$ in **e**–**g** are highlighted in solid grey.



Extended Data Fig. 6 | Energy relaxation and decoherence. Measurement of the energy relaxation time T_1 and spin-echo coherence time T_2 at an external flux bias point close to the half-flux quantum. The measured time traces are fitted with decaying exponents. In the spin-echo sequence, the refocusing π rotation was applied around the axis perpendicular to the axis of the two π/2 rotations.



Extended Data Fig. 7 | Dephasing limit. Estimated dephasing time T_φ due to the first-order sensitivity of the qubit transition frequency f_{01} to the flux noise. Here, we use $1/T_\varphi = 2\pi \frac{\partial f_{01}}{\partial \Phi} A\sqrt{\ln 2}$, where Φ is the total magnetic flux through the loop, and assume typical flux noise amplitude[26] of $A \approx 1.8\times10^{-6}(h/2e)$.

A submicrometre silicon-on-insulator resonator for ultrasound detection

https://doi.org/10.1038/s41586-020-2685-y

Received: 26 July 2018

Accepted: 3 July 2020

Rami Shnaiderman[1,2]✉, **Georg Wissmeyer**[1,2,3], **Okan Ülgen**[1,2,3], **Qutaiba Mustafa**[1,2], **Andriy Chmyrov**[1,2] & **Vasilis Ntziachristos**[1,2]✉

Ultrasound detectors use high-frequency sound waves to image objects and measure distances, but the resolution of these readings is limited by the physical dimensions of the detecting element. Point-like broadband ultrasound detection can greatly increase the resolution of ultrasonography and optoacoustic (photoacoustic) imaging[1,2], but current ultrasound detectors, such as those used for medical imaging, cannot be miniaturized sufficiently. Piezoelectric transducers lose sensitivity quadratically with size reduction[3], and optical microring resonators[4] and Fabry–Pérot etalons[5] cannot adequately confine light to dimensions smaller than about 50 micrometres. Micromachining methods have been used to generate arrays of capacitive[6] and piezoelectric[7] transducers, but with bandwidths of only a few megahertz and dimensions exceeding 70 micrometres. Here we use the widely available silicon-on-insulator technology to develop a miniaturized ultrasound detector, with a sensing area of only 220 nanometres by 500 nanometres. The silicon-on-insulator-based optical resonator design provides per-area sensitivity that is 1,000 times higher than that of microring resonators and 100,000,000 times better than that of piezoelectric detectors. Our design also enables an ultrawide detection bandwidth, reaching 230 megahertz at −6 decibels. In addition to making the detectors suitable for manufacture in very dense arrays, we show that the submicrometre sensing area enables super-resolution detection and imaging performance. We demonstrate imaging of features 50 times smaller than the wavelength of ultrasound detected. Our detector enables ultra-miniaturization of ultrasound readings, enabling ultrasound imaging at a resolution comparable to that achieved with optical microscopy, and potentially enabling the development of very dense ultrasound arrays on a silicon chip.

Ultrasound detection using optical methods offers a fundamental advantage over piezoelectric detection because the detectors can be miniaturized without sacrificing sensitivity[3]. A common example is optical interferometry using a π-shifted Bragg-grating etalon embedded in a fibre waveguide[8,9]. In this configuration, ultrasound waves perturb the length and refractive index of the optical cavity established between two Bragg gratings, changing its resonance characteristics. However, the large sensing length (100–300 μm)[9,10] and narrow bandwidth (10–30 MHz)[9,10] do not allow point-like detection, limiting resolution and miniaturization potential. Other resonator designs include polymer microrings[4,11] or Fabry–Pérot etalons[5,12], but light confinement requirements dictate sizes in the tens of micrometres, limiting further miniaturization.

We introduce a concept for ultrasound detection, based on the highly scalable silicon-on-insulator (SOI) platform, that exploits the high-throughput fabrication techniques that are widely used in the semiconductor industry. Using this technology, we designed a point-like silicon waveguide–etalon detector (SWED). With dimensions of 220 nm × 500 nm, it is four orders of magnitude smaller than the smallest polymer microring detectors[4] and an order of magnitude smaller than the diameter of cells and blood capillaries. We show that our concept enables greatly improved ultrasound detection, despite the miniaturization achieved.

The SWED (Fig. 1a, b) contains a single continuous silicon waveguide divided into four sections: an Ag layer, a spacer, a cavity and a Bragg grating (Fig. 1b). A metallic reflective layer approximately 200 nm thick (Fig. 1a, b, 'Ag') was deposited onto the polished end facet of the waveguide, followed by a spacer section consisting of an ultrashort Bragg grating with a varying length of around 3–25 μm (Fig. 1a, b, 'spacer'). The metallic reflective layer and the spacer form the first optical mirror of the etalon. Use of an ultrathin metallic layer places the optical cavity close to the end facet of the waveguide and enables ultrasound detection through its cross-section without substantial attenuation. A 320-nm-long waveguide segment lies adjacent to the spacer, forming the etalon cavity (Fig. 1a, b, 'cavity'). The other end of the cavity is terminated by a second optical mirror, made of a 125-μm-long Bragg grating (Fig. 1a, b, 'Bragg grating'). The length of the cavity was designed so that the acquired optical round-trip phase shift is π at the resonance

---

[1] Chair of Biological Imaging and TranslaTUM, Technische Universität München, Munich, Germany. [2] Institute of Biological and Medical Imaging, Helmholtz Zentrum München, Neuherberg, Germany. [3] These authors contributed equally: Georg Wissmeyer, Okan Ülgen. ✉e-mail: rami.shnaiderman@tum.de; v.ntziachristos@tum.de

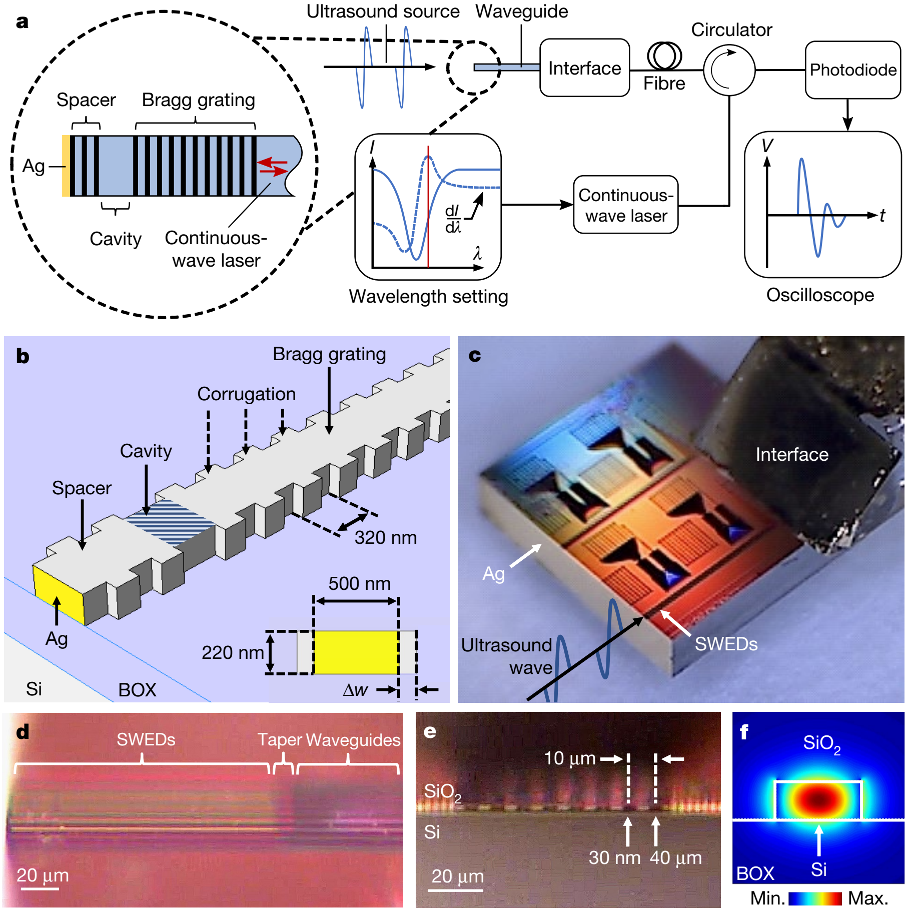


Fig. 1 | Design and operating principles of the SWED. a, SWED read-out system. A continuous-wave laser pumps light into the SWED. Incident ultrasound on the SWED induces variations to the reflected optical intensity (*I*) from the SWED, which is diverted to a photodetector by the fibre circulator. The intensity variations are recorded by an oscilloscope as voltage (*V*) variations as a function of time (*t*). The wavelength (λ) of the continuous-wave laser is tuned off-resonance to the point of the steepest slope. The dashed curve in 'wavelength setting' indicates the slope of the resonance as a function of wavelength ($dI/d\lambda$); the red vertical line indicates the wavelength the continuous-wave laser. **b**, Schematic of a single SWED. The corrugation depth on the sides of the Bragg grating is defined as Δw (inset). BOX indicates the silicon-oxide substrate of the silicon waveguide. **c**, Photograph of the SOI chip with an array of eight SWEDs facing the ultrasound source. **d**, Brightfield micrograph of the SOI chip taken perpendicularly to the optical axis of the SWEDs. Scale bar, 20 µm. **e**, Brightfield micrograph of the SOI chip taken in the direction of the optical axis of the SWED, before the application of the Ag coating. Scale bar, 20 µm. **f**, Normalized profile of the horizontal component of the electric field (colour scale) over an area of 1 µm × 1 µm. The white lines indicate the boundaries of the waveguide and the silicon-oxide substrate.

wavelength of the etalon. The Bragg gratings were constructed by adding lateral corrugation (Fig. 1b, 'corrugation') to the waveguide, with a corrugation depth of Δw (Fig. 1b, inset), periodicity of 320 nm and duty cycle of 50%.

We manufactured four SOI chips, each measuring 3 mm × 3 mm × 0.8 mm (Fig. 1c), with spacer lengths of 26 µm, 14 µm, 9 µm and 3.5 µm. Each chip comprises eight SWEDs, aligned next to each other with a pitch of 10 µm (Fig. 1e). The measured optical and acoustic properties of the SWEDs are summarized in Extended Data Table 1.

A brightfield micrograph of one such chip, taken perpendicularly to the optical axis of the SWEDs (Fig. 1d), depicts the eight SWEDs, each connected through a 15-µm adiabatic taper to eight silicon waveguides, leading to an interface (Fig. 1c; see Methods). The interface is connected on its other side (see Fig. 1a) to an array of eight single-mode, polarization-maintaining fibres that connect each of the SWEDs to the circulator, continuous-wave laser and photodiode. A brightfield micrograph of the chip, taken in the direction of the optical axis of the SWEDs (Fig. 1e) and obtained before the application of the Ag coating, depicts the cross-sections of the eight SWEDs. The SWEDs were designed with different corrugation depths Δw; Fig. 1e highlights SWEDs manufactured with $\Delta w = 30$ nm and $\Delta w = 40$ nm.

The cross-section of the SWED is 220 nm in height and 500 nm wide (Fig. 1b), and supports a well-confined optical transverse electric mode (Fig. 1f). The SOI platform offers particularly high-index contrast ($\Delta n \approx 2.5$) between the cladding and the cavity materials of the SWED, enabling highly effective light confinement across the waveguide, at optical subwavelength cross-sections[13,14]. With a cavity length of only 320 nm, the SWED has a single, very narrow resonance at near-infrared wavelengths (Fig. 2a).

To detect ultrasound waves, the continuous-wave laser pumps light into the cavity of the SWED. To enhance sensitivity, the laser is tuned off-resonance (Fig. 1a, 'wavelength setting') and the polarization is maintained in the transverse electric orientation by the polarization-maintaining fibres (Fig. 1a, 'fibre'). Off-resonance tuning places the pump wavelength at the maximum slope of the resonance curve of the etalon (Fig. 1a, dashed curve in 'wavelength setting'), ensuring that the optical phase variation in response to incident ultrasound waves (Fig. 1a, 'ultrasound waves') is amplified by the SWED (see Supplementary Information). The reflected light modulated by the ultrasound waves is detected by a photodiode and recorded by an oscilloscope (Fig. 1a).

To characterize the detection bandwidth, we focused a pulsed laser beam onto a 200-nm-thick gold film, generating a broadband

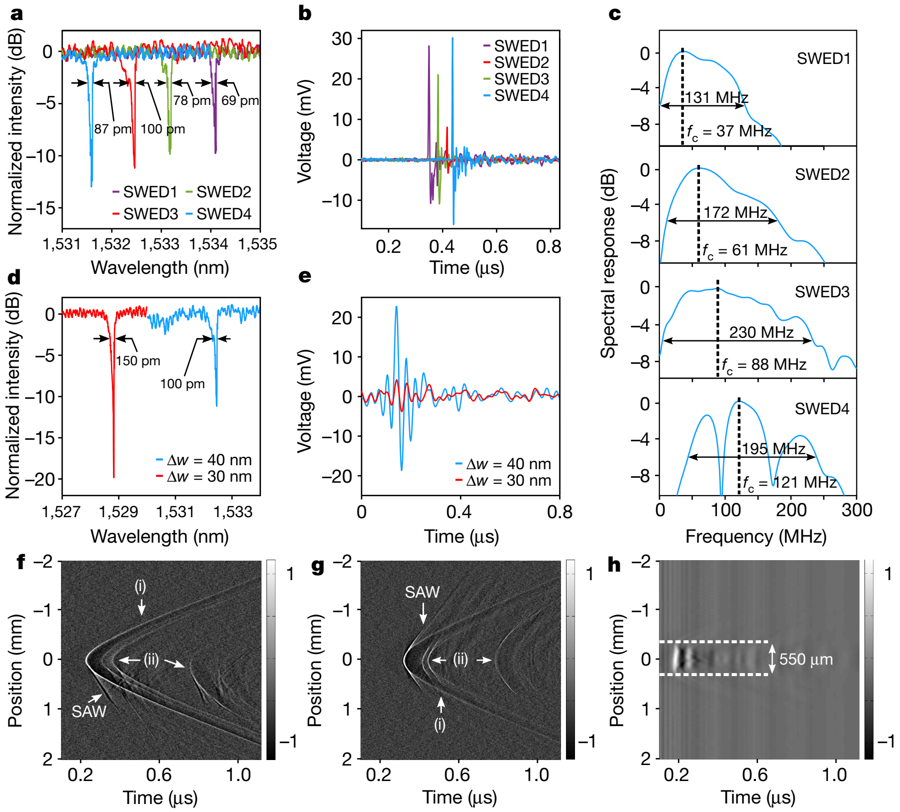


Fig. 2 | Characterization of the SWEDs. a, Optical reflection spectra of four SWEDs with varying spacer lengths (as labelled; see Extended Data Table 1). **b**, Temporal responses of the SWEDs in **a** following exposure to a broadband ultrasound point source. The SWEDs were placed at the same distance from the source, but each succeeding signal is shifted by 33 ns for clear presentation. **c**, Spectral responses of SWEDs to the signals in **b**. The detection bandwidths (arrows) and central detection frequencies f_c (vertical dashed lines) are indicated for each SWED. **d**, Optical reflection spectra of SWED3$_{40nm}$ (blue) and SWED3$_{30nm}$ (red). The two SWEDs differ in the corrugation depth Δw. **e**, Temporal response of the two SWEDs in **d** following exposure to an ultrasound source. **f**, **g**, Spatial responses of SWED3$_{40nm}$ acquired by scanning the SWED linearly over a broadband ultrasound point source. The scanning path was either parallel to the short dimension of the chip facet (**f**) or perpendicular to the long dimension of the chip facet (**g**). Pressure amplitude is depicted on a continuous greyscale; (i) indicates longitudinal waves; SAW indicates surface acoustic waves; (ii) indicates reflections from the sample holder. **h**, Spatial response of a needle hydrophone with diameter of 0.5 mm acquired by scanning the hydrophone over a broadband ultrasound point source. The point source is smeared to around 550 µm, close to the Φ_d = 0.5 mm of the hydrophone.

ultrasound point source. SWED measurements (Fig. 2b, c) of the ultrasound waves generated, using water as a coupling medium, revealed that changes in spacer length can tune the central detection frequency (f_c) and shift the detection bandwidth. We found that f_c increases with decreasing spacer length (Extended Data Table 1), possibly owning to enhanced detection of shorter wavelengths when the optical cavity, the most sensitive section of the SWED, is located closer to the chip facet. Figure 2c shows a detection bandwidth as large as 230 MHz, owing to the subwavelength sensing area of the SWED.

We investigated the sensitivity of the SWED using a needle hydrophone calibrated in the range 5–30 MHz. Ultrasonic signals were generated by focusing a pulsed laser onto a 125-µm-thick black vinyl. The ultrasound source was positioned at a fixed distance, first in front of the hydrophone and then in front of the SWED, using water for acoustic coupling. The SWED with a spacer length of 26 µm (SWED1) has a bandwidth that most closely overlaps with the calibrated bandwidth of the hydrophone, and was therefore selected for characterization. The noise-equivalent pressure (NEP) for SWED1 was determined to be 45 Pa (9 mPa Hz$^{-1/2}$) over a 25-MHz bandwidth around the central frequency f_c = 37 MHz (see Methods).

Because the bandwidths of the other SWEDs do not overlap with the calibrated bandwidth of the hydrophone, we deduced their sensitivity relative to SWED1. Figure 2b shows the response of the SWEDs to a broadband ultrasound point source with approximately flat bandwidth. SWED2 (spacer length of 14 µm) and SWED4 (spacer length of 3.5 µm) exhibit similar sensitivity to SWED1 (Fig. 2b). The sensitivity of SWED3 (spacer length of 9 µm) is less than that of SWED1 by a factor of 3.3, as a result of its cavity having a lower Q-factor (a measure of the relative linewidth of the resonances; Extended Data Table 1), probably owing to inadequate metallic coating.

To demonstrate the ability to read from consecutive SWEDs on a single chip, we serially interrogated adjacent detectors on the SWED3 chip. These measurements enabled us to study the effect of the corrugation depth on the performance of the SWED. Figure 2d shows the resonances for SWED3 with Δw = 40 nm (SWED3$_{40nm}$) and Δw = 30 nm (SWED3$_{30nm}$). We found that the deeper the corrugation, the higher the Q-factor (Extended Data Table 1), owing to the increased reflectivity of the Bragg gratings; the sensitivity of SWED3$_{40nm}$ was found to be roughly four times better than that of SWED3$_{30nm}$ (Fig. 2e). Changes in the corrugation depths affect the sensitivity of the SWED more strongly than do changes in the spacer lengths (compare Fig. 2b, e).

We characterized the spatial response of the SWEDs by scanning SWED3$_{40nm}$ over the ultrasound point source used for bandwidth determination. The time trace of the pressure amplitude was recorded for each scanning step (B-scan), resulting in characteristic curved profiles for the detected pulses (Fig. 2f, g, profile (i)). A linear fit of profile (i) revealed a sound velocity of approximately 1,526 m s^{-1}—in good agreement with the velocity of longitudinal acoustic waves in water—indicating a direct propagation path from the source to the sensing area of the SWED. Acoustic reflections between the SWED and the sample holder were also observed (Fig. 2f, g, profile (ii)). We also observed pulses with a dominant negative dip, which propagate at much faster velocities and correspond to surface acoustic waves, that is, Rayleigh waves and bulk shear waves[15] (Fig. 2f, g, 'SAW').

The acceptance angle for longitudinal ultrasonic waves was calculated from profile (i) in Fig. 2f to be 148°, which corresponds to an acoustic numerical aperture (NA) of 0.96. This value is close to the value of NA = 1 for a theoretical point detector, confirming the ultrasmall dimensions of the sensing area of the SWED relative to the wavelength of the ultrasound waves that were detected. This performance could lead to higher lateral resolution compared to what is achieved with conventional detectors. The lateral resolution (R_L) of an unfocused detector of size Φ_d can be written as[16]

$$R_\mathrm{L} = [(R_\mathrm{A})^2 + (\Phi_\mathrm{d})^2]^{1/2}, \tag{1}$$

where $R_\mathrm{A} \approx 0.8 v_\mathrm{s}/f_\text{cut-off}$ is the axial resolution, v_s is the speed of sound and $f_\text{cut-off}$ is the cut-off frequency of the detector. According to equation (1), a point detector with negligible Φ_d yields $R_\mathrm{L} \approx R_\mathrm{A}$, that is, isometric resolution. Conversely, strong lateral smearing of the ultrasound point source is observed (Fig. 2h) when scanning over it with a needle hydrophone with non-negligible $\Phi_\mathrm{d} = 0.5$ mm.

The point-like nature of the SWED opens up new possibilities for ultrasonic measurements. In analogy to near-field scanning optical microscopy[17,18], the subwavelength dimensions of the SWED could enable super-resolution investigations in ultrasonics and optoacoustics. To explore this possibility, we performed B-scans using SWED4 over a polystyrene suture of 30 μm diameter in the acoustic far field (Fig. 3a; distance of 0.2 mm or 6.7 wavelengths λ) and in the near field (Fig. 3c; distance of 20 μm or 0.7λ). Inspection of the near-field measurements (Fig. 3d) reveals a sharp transition between the acoustic signals originating in the near field and the far field. This transition distinguishes the evanescent waves from the far-field propagating wavefront. In the far-field scan, only the propagating wavefront is detected (Fig. 3b). Owing to the point-like nature of the SWED, it is possible to accurately resolve the profile of the suture from the near-field measurement, without the need for image reconstruction—that is, by observing only the raw data (B-scan) collected.

The performance seen in Fig. 3d highlights the effect—similar to that observed for near-field scanning optical microscopy—of sensing the high-spatial-frequency components in the near field with an aperture smaller than the imaging wavelength. To confirm this effect, using the same pulsed laser and SWED4, we performed optoacoustic imaging of a marking in the shape of the number '6' on a 1951 US Air Force (USAF) resolution test target. To enhance the images obtained in the far field (Fig. 3e; 6.2λ distance), raw data from the B-scan was reconstructed (see Methods) and shown as a maximum intensity projection. Substantial blurring is still seen, despite the improvement in the image due to inversion (Fig. 3f). By contrast, raw images (without reconstruction) obtained in the near field (Fig. 3g; 0.46λ distance) are sharp and high-resolution, with no blurring (Fig. 3h). Figure 3i, j shows the profiles of the near-field and far-field images along the white dashed lines in Fig. 3f, demonstrating the improvement achieved in the near field owing to the point-like properties of the SWED.

To confirm the subwavelength resolution enabled by the small dimensions of the SWED, we benchmarked the near-field imaging with SWED4 against a high-end optical confocal microscope with a 60× water immersion objective and NA = 1.2 (roughly 230-nm resolution; see Methods). We used the optical microscope (Fig. 3k) and the SWED (Fig. 3l; 0.43λ distance) to image a thin metallic hexagonal grating used to calibrate electron microscopes. Enlarging the section enclosed by the white dashed box in Fig. 3l shows that the SWED resolved the finest features of the sample without adding artefacts (Fig. 3m). A detailed comparison of the profiles taken along the white dashed lines in Fig. 3k, m reveals that both the optical microscope and SWED4 produce images of similar resolution and contrast. The width of the hexagon ridge (Fig. 3n) was measured to be about 9 μm for both the optical microscope and the SWED, in close agreement with the actual, specified dimensions (8 μm). To quantify the lateral SWED resolution precisely, we measured the edge-spread function of an edge on the same 1951 USAF resolution test target. SWED4 resolved the edge with virtually the same edge-spread function as the confocal microscope (Fig. 3o). Using Richardson–Lucy deconvolution and the optical point-spread function of the confocal microscope, we were able to estimate the point-spread function of the SWED (see Methods) to be 650 nm (Fig. 3p). This value is close to the manufactured sensing area of the detector, confirming the capability of submicrometre detection of sound.

These findings highlight an interesting property of the SWED. Even though imaging of the edge was performed at 46 MHz (the strongest frequency emitted from the target), with expected diffraction blurring of the order of the acoustic wavelength (32.6 μm), the target is imaged with a resolution of 0.65 μm. Thus, the SWED demonstrates super-resolution performance, being able to resolve features 50 times smaller than the wavelength.

The investigations illustrated in Fig. 3 were performed in the near field; however, we postulated that even in the far field the ultrahigh bandwidth of the SWED would allow observations at acoustic resolutions not previously demonstrated. By imaging acoustic interference, we show that it is possible to resolve patterns generated at frequencies of tens to hundreds of megahertz. Imaging acoustic interference was previously only possible using bulky structures designed to transmit information from the near field to the far field for a single frequency in the kilohertz range[19–21]. The interfering waves that we measured were generated by illuminating a 200-nm-thick gold film with a beam about 260 μm in diameter generated by a pulsed laser. A B-scan with $\mathrm{SWED3_{40nm}}$ over the gold film, along the diameter of the beam, measured a planar wave and a focal spot formed by the interference of waves travelling diagonally from the edges of the illuminating beam (Fig. 4a). The spot measures 20 μm laterally and contains several distinct frequency bands (I–IV; Fig. 4b). Inspecting each frequency band separately by applying a band-pass filter to the raw data reveals complex fringe patterns (Fig. 4c–f). The responses of the SWED along the profiles indicated by the white arrows in Fig. 4c–f are depicted in Fig. 4g–j. The full-width at half-maximum (FWHM) of the narrowest fringe in each profile, relative to the noise floor, shows that at the lowest frequency band the fringe is roughly 50 μm wide, owing to the long wavelengths and contributions from beating frequencies, whereas in the highest band the fringe narrows to around 5 μm. The latter value is in good agreement with the theoretical diffraction-limited lateral resolution of $\mathrm{SWED3_{40nm}}$ (equation (1)). This finding demonstrates the ability of the SWED to resolve fine acoustic patterns, enabled by its submicrometre sensing area. In comparison to hydrophone measurements[21], the measurement using the SWED are finer by a factor of at least 800.

We developed a miniaturized ultrasound detector, with a sensing area at least 450 times smaller than that of π-shifted Bragg grating etalons[10] and 26,000 times smaller than that of polymer microrings[4]. With dimensions up to 200 times smaller than the acoustic wavelengths detected, the SWED satisfies the definition of a true point detector. We used the SWED to perform SOI-based imaging, demonstrating $\lambda/50$ resolution in the near field. This resolution is comparable to that of optical microscopes (up to about $\lambda/500$ when selecting detection in the lower ultrasound frequencies to exemplify the performance). As a result, the SWED was able to resolve high-frequency acoustic interference fringes as narrow as around 5 μm in the far field, which is not possible with conventional detectors[22].

We compared the performance of the SWED with other detectors, using the product of the NEP and the sensing area as a figure of merit. This product has been reported to be 1.58×10^{-2} mPa mm^2 Hz$^{-1/2}$ (NEP = 5.6 mPa Hz$^{-1/2}$) for polymer microrings used in optoacoustic microscopy[23] and 509 mPa mm^2 Hz$^{-1/2}$ (NEP = 450 mPa Hz$^{-1/2}$) for a

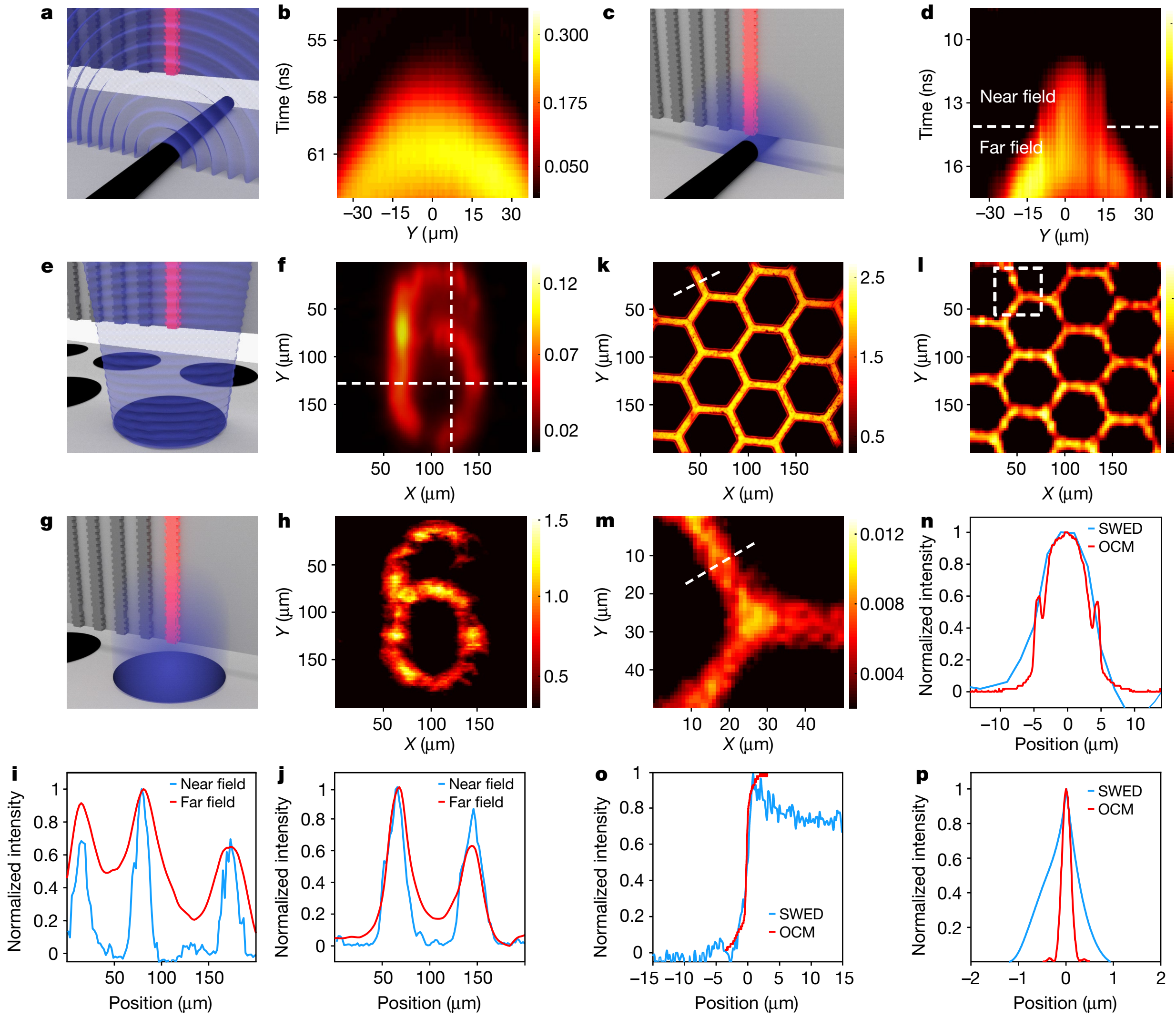


Fig. 3 | Reflection-mode far-field and near-field optoacoustic imaging. **a**, Illustration of far-field imaging of a suture (black). The SWED with the laser light pumped into it is indicated in pink; other SWEDS, which are not interrogated during this experiment, are visible to the left of the pink SWED. The propagating acoustic waves in the far field are shown in purple. **b**, A far-field B-scan of the suture with SWED4. The colour scale indicates the amplitude of the detected ultrasonic signals. **c**, Illustration of near-field imaging of a suture. The evanescent acoustic waves in the near field are shown in purple. **d**, A near-field B-scan of the suture with SWED4. The white dashed line divides the signals detected in the near field and the far-field. The transition between the imaging regimes is sharp, resulting in a roughly 3 ns delay in the arrival time for the far-field signals. **e**, Illustration of far-field imaging of a marking on a 1951 USAF resolution test target. The propagating acoustic waves in the far field are shown in purple. **f**, Image formed by a maximum intensity projection and reconstruction of the marking acquired with SWED4 in the far field. **g**, Illustration of near-field imaging of the same marking. The evanescent acoustic waves in the near field are shown in purple. **h**, Image formed by maximum intensity projection of only the near-field signals acquired with SWED4. **i**, **j**, Comparison of the normalized intensity in **f** (red, far field) and **h** (blue, near field) along the vertical (**i**) and horizontal (**j**) dashed white lines in **f**. **k**, Image of a gold hexagonal grid acquired by an optical confocal microscope (OCM; 60× objective, numerical aperture of 1.2). **l**, Image of the same hexagonal grid acquired by SWED4 in the near field. **m**, Enlargement of the section enclosed by the white dashed box in **l**. **n**, Comparison of the normalized intensity along the dashed lines in **k** (red, OCM) and **m** (blue, SWED). **o**, Edge-spread functions of a straight edge on the resolution test target acquired with SWED4 in the near field (blue) and the confocal microscope (red). **p**, Point-spread functions of SWED4 (blue) and the confocal microscope (red).

miniaturized piezoelectric transducer used for intravascular ultrasound imaging[24]. These values correspond to a sensitivity three and eight orders of magnitude, respectively, lower than that achieved by the SWED (9.9×10^{-7} mPa mm^2 Hz$^{-1/2}$; NEP = 9 mPa Hz$^{-1/2}$). This marked sensitivity improvement is attributed to the high spatial confinement of light in the SOI platform. The sensitivity could be further improved by increasing the Q-factor of the SWED; this could be achieved by increasing the reflectivity of the Bragg grating and the metallic coating, and by reducing the optical losses by smoothing waveguide side walls[25] and using rib waveguide geometries[26]. Acoustical matching between the impedances of silicon (around 20 MRayl) and biological samples (around 1.4 MRayl) via the application of thin layers will also improve sensitivity.

We demonstrated the design flexibility of our detector and the SOI platform by producing several SWEDs with different detection bandwidths. We achieved a bandwidth as large as 230 MHz and were able to shift it on-demand by more than 80 MHz without reducing the relative sensitivity of the SWED in each frequency band. We also demonstrated a line array of eight SWEDs with a one-dimensional detection density of 100 detectors per millimetre and a two-dimensional detection density of 125 detectors per square millimetre. This result is an improvement in density by orders of magnitude compared to state-of-the-art

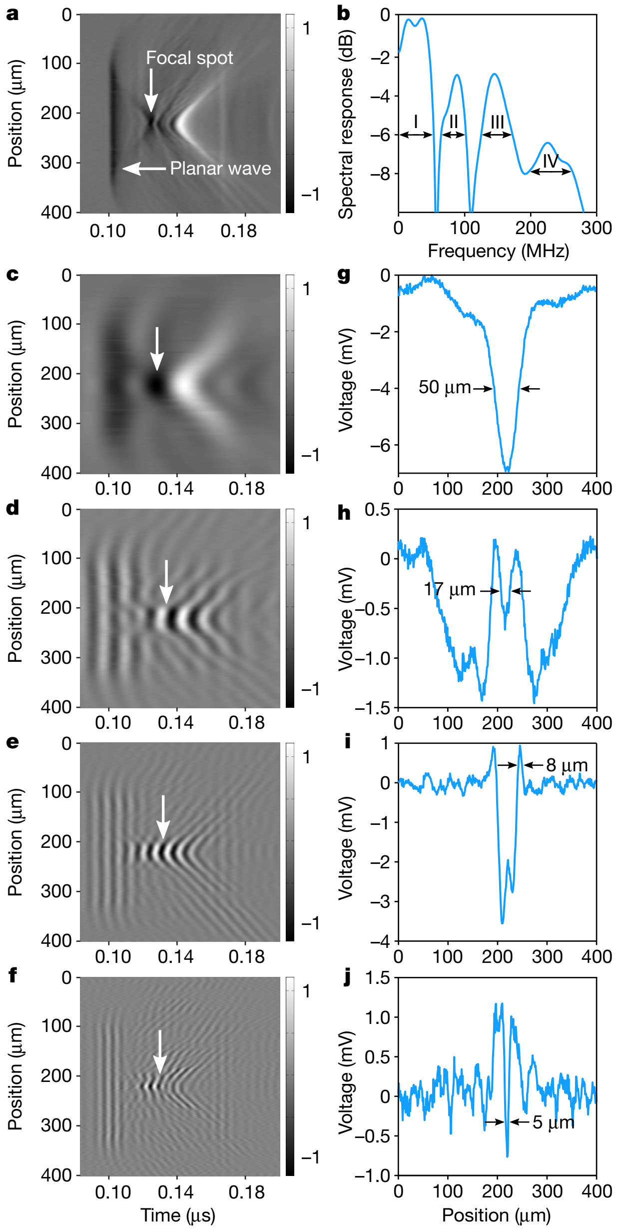


Fig. 4 | Imaging of acoustic broadband interference. a, B-scan with SWED3$_{40nm}$ over the diameter of the laser beam incident on the gold film. A planer wave is emitted from the gold plate; the focal spot forms as a result of the interference of waves emitted from the edges of the excitation beam. **b**, Frequency content of the focal spot in **a**. I–IV indicate distinct frequency bands. **c**–**f**, Interference patterns observed in frequency bands I–IV, respectively. **g**–**j**, Response of the SWED along the profiles indicated by the white arrows in **c**–**f**, respectively. The FWHM of the narrowest fringe (relative to the average noise floor) is labelled in each panel.

piezoelectric arrays (9.5 detectors per square millimetre)[22] and arrays of capacitive micromachined ultrasound transducers (2.5 detectors per square millimetre)[27]. The high-throughput, scalable semiconductor fabrication allowed by the SOI technology means that the detectors can be mass-produced. In comparison, arrays of polymer microrings are difficult to manufacture and, owing to the poor light confinement of the polymer platform, may result in large device footprints and limited integration density. Likewise, ultrasound detectors based on focused light beams are cumbersome and expensive, as they generally require high-end objectives[28] and possibly require elaborate optical systems for signal detection[29,30]. Therefore, they are not well suited for miniaturized or highly disseminated applications. In addition, they have not yet achieved the resolution or bandwidth demonstrated by SWEDs[12]. Our technology could be scaled to densities of more than 1,000 detectors per square millimetre, as the inter-SWED distances and silicon wafer thickness can be reduced using currently available manufacturing processes. Two-dimensional arrays could be manufactured by stacking chips on top of each other or by using multilayer manufacturing processes[31].

Nevertheless, large-scale array multiplexing is challenging for optical resonator devices. One of the challenges of many photonic platforms is the strong dependence of the resonance on device size. This cannot be precisely controlled owing to process-dependent variations[32], making the interrogation of a large number of resonators with a single continuous-wave laser difficult. The SOI platform offers advantages over other photonic approaches, which may lead to practical solutions for multiplexing. On-chip photodetectors[33] and tuneable Mach–Zander interferometers[34], which are fully compatible with the manufacturing process for SWEDs, could be placed alongside the SWEDs and could greatly reduce the costs of multiplexing when using pulse interferometry[35]. The intrachip uniformity of the SOI platform[36] and the large integration density of the SWEDs (compared to, for example, polymer-based platforms) could also be used to reduce resonance-wavelength-variability and laser-tuning requirements, enabling multiplexing of several tens of SWEDs with a single continuous-wave laser.

The high bandwidth and submicrometre, point-like aperture of the SWED leads to detection performance that substantially improves the axial and lateral resolution in ultrasound and optoacoustic imaging. Increases in bandwidth are essential for achieving high tomographic resolution[16], and the submicrometre aperture of the SWED could lead to greater lateral resolution, potentially setting a new standard for non-invasive ultrasound and optoacoustic imaging[37–39]. The combination of SWEDs with SOI-based integrated biosensors[40] or on-chip microscopes[41] could lead to powerful tools for basic research and diagnostics.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2685-y.

1. Ntziachristos, V. Going deeper than microscopy: the optical imaging frontier in biology. *Nat. Methods* **7**, 603–614 (2010).
2. Wang, L. V. & Yao, J. A practical guide to photoacoustic tomography in the life sciences. *Nat. Methods* **13**, 627–638 (2016).
3. Wissmeyer, G., Pleitez, M. A., Rosenthal, A. & Ntziachristos, V. Looking at sound: optoacoustics with all-optical ultrasound detection. *Light Sci. Appl.* **7**, 53 (2018).
4. Zhang, C., Chen, S. L., Ling, T. & Guo, L. J. Review of imprinted polymer microrings as ultrasound detectors: design, fabrication, and characterization. *IEEE Sens. J.* **15**, 3241–3248 (2015).
5. Guggenheim, J. A. et al. Microresonators for ultrasound sensing. *Nat. Photon.* **11**, 714–719 (2017).
6. Khuri-Yakub, B. T. & Oralkan, Ö. Capacitive micromachined ultrasonic transducers for medical imaging and therapy. *J. Micromech. Microeng.* **21**, 54004–54014 (2011).
7. Qiu, Y. et al. Piezoelectric micromachined ultrasound transducer (PMUT) arrays for integrated sensing, actuation and imaging. *Sensors* **15**, 8020–8041 (2015).
8. Shnaiderman, R. et al. Fiber interferometer for hybrid optical and optoacoustic intravital microscopy. *Optica* **4**, 1180–1187 (2017).
9. Rosenthal, A., Razansky, D. & Ntziachristos, V. High-sensitivity compact ultrasonic detector based on a pi-phase-shifted fiber Bragg grating. *Opt. Lett.* **36**, 1833–1835 (2011).
10. Rosenthal, A. et al. Embedded ultrasound sensor in a silicon-on-insulator photonic platform. *Appl. Phys. Lett.* **104**, 021116 (2014).
11. Dong, B. et al. Isometric multimodal photoacoustic microscopy based on optically transparent micro-ring ultrasonic detection. *Optica* **2**, 169–176 (2015).
12. Jathoul, A. P. et al. Deep in vivo photoacoustic imaging of mammalian tissues using a tyrosinase-based genetic reporter. *Nat. Photon.* **9**, 239–246 (2015).
13. Wang, X. et al. Narrow-band waveguide Bragg gratings on SOI wafers with CMOS-compatible fabrication process. *Opt. Express* **20**, 15547–15558 (2012).
14. Lockwood, D. J. & Pavesi, L. (eds) *Silicon Photonics II. Components and Integration* Ch. 1 (Springer, 2011).

15. Royer, D. & Dieulesaint, E. *Elastic Waves in Solids I. Free and Guided Propagation* Ch. 5 (Springer, 2000).
16. Xu, M. & Wang, L. V. Analytic explanation of spatial resolution related to bandwidth and detector aperture size in thermoacoustic or photoacoustic reconstruction. *Phys. Rev. E* **67**, 056605 (2003).
17. Ash, E. A. & Nicholls, G. Super-resolution aperture scanning microscope. *Nature* **237**, 510–512 (1972).
18. Harootunian, A., Betzig, E., Isaacson, M. & Lewis, A. Super-resolution fluorescence near-field scanning optical microscopy. *Appl. Phys. Lett*. **49**, 674–676 (1986).
19. Huang, F. M. & Zheludev, N. I. Super-resolution without evanescent waves. *Nano Lett*. **9**, 1249–1254 (2009).
20. Li, J., Fok, L., Yin, X., Bartal, G. & Zhang, X. Experimental demonstration of an acoustic magnifying hyperlens. *Nat. Mater*. **8**, 931–934 (2009).
21. Zhu, J. et al. A holey-structured metamaterial for acoustic deep-subwavelength imaging. *Nat. Phys*. **7**, 52–55 (2011).
22. Omar, M., Aguirre, J. & Ntziachristos, V. Optoacoustic mesoscopy for biomedicine. *Nat. Biomed. Eng*. **3**, 354–370 (2019).
23. Li, H., Dong, B., Zhang, Z., Zhang, H. F. & Sun, C. A transparent broadband ultrasonic detector based on an optical micro-ring resonator for photoacoustic microscopy. *Sci. Rep*. **4**, 4496 (2015).
24. Rosenthal, A. et al. Sensitive interferometric detection of ultrasound for minimally invasive clinical imaging applications. *Laser Photonics Rev*. **8**, 450–457 (2014).
25. Gao, F., Wang, Y., Cao, G., Jia, X. & Zhang, F. Reduction of sidewall roughness in silicon-on-insulator rib waveguides. *Appl. Surf. Sci*. **252**, 5071–5075 (2006).
26. Bogaerts, W. & Selvaraja, S. K. Compact single-mode silicon hybrid rib/strip waveguide with adiabatic bends. *IEEE Photonics J*. **3**, 422–432 (2011).
27. Huang, X. et al. A high-frequency transimpedance amplifier for CMOS integrated 2D CMUT array towards 3D ultrasound imaging. In *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)* (eds Hielscher, A. & Liu, S.) 101–104 (IEEE, 2013).
28. Hajireza, P., Shi, W., Bell, K., Paproski, R. J. & Zemp, R. J. Non-interferometric photoacoustic remote sensing microscopy. *Light Sci. Appl*. **6**, e16278 (2017).
29. Pelivanov, I., Buma, T., Xia, J., Wei, C. & Donnell, M. O. A new fiber-optic non-contact compact laser-ultrasound scanner for fast non-destructive testing and evaluation of aircraft composites. *J. Appl. Phys*. **115**, 113105 (2014).
30. Pelivanov, I. et al. High resolution imaging of impacted CFRP composites with a fiber-optic laser-ultrasound scanner. *Photoacoustics* **4**, 55–64 (2016).
31. Sacher, W. D. et al. Monolithically integrated multilayer silicon nitride-on-silicon waveguide platforms for 3-D photonic circuits and devices. *Proc. IEEE* **106**, 2232–2245 (2018).
32. Krishnamoorthy, A. V. et al. Exploiting CMOS manufacturing to reduce tuning requirements for resonant optical devices. *IEEE Photonics J*. **3**, 567–579 (2011).
33. Michel, J., Liu, J. & Kimerling, L. C. High-performance Ge-on-Si photodetectors. *Nat. Photon*. **4**, 527–534 (2010).
34. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photon*. **11**, 441–446 (2017).
35. Rosenthal, A., Razansky, D. & Ntziachristos, V. Wideband optical sensing using pulse interferometry. *Opt. Express* **20**, 19016–19029 (2012).
36. Selvaraja, S. K., Bogaerts, W., Dumon, P., Van Thourhout, D. & Baets, R. Subnanometer linewidth uniformity in silicon nanophotonic waveguide devices using CMOS fabrication technology. *IEEE J. Sel. Top. Quantum Electron*. **16**, 316–324 (2010).
37. Wang, L. V. & Hu, S. Photoacoustic tomography: in vivo imaging from organelles to organs. *Science* **335**, 1458–1462 (2012).
38. Tzoumas, S. et al. Eigenspectra optoacoustic tomography achieves quantitative blood oxygenation imaging deep in tissues. *Nat. Commun*. **7**, 12121 (2016).
39. Taruttis, A. & Ntziachristos, V. Advances in real-time multispectral optoacoustic imaging and its applications. *Nat. Photon*. **9**, 219–227 (2015).
40. Sang, S. et al. Progress of new label-free techniques for biosensors: a review. *Crit. Rev. Biotechnol*. **36**, 465–481 (2016).
41. Diekmann, R. et al. Chip-based wide field-of-view nanoscopy. *Nat. Photon*. **11**, 322–328 (2017).

Methods

Device fabrication

The chip layout was designed using Optodesigner software (PhoeniX Software). It consists of several components: waveguides, grating-couplers, Bragg gratings, tapers and arrayed waveguide gratings (the arrayed gratings were not used in this study and are not discussed here). The chip was fabricated at the Interuniversity Microelectronics Centre (IMEC) through the ePIXfab Consortium Service on an SOI wafer with silicon orientation of (100). The main fabrication techniques included ultraviolet-lithography on a standard I-line resist followed by a two-step etch process of the silicon, involving a shallow etch (70 nm) and a deep etch (220 nm). The components were embedded between 2 μm of SiO_2 back oxide (Fig. 1c, 'BOX') and 1.25 μm of SiO_2 cladding. The waveguides were designed to be single-mode with a cross-section of 220 nm in height and 450 nm or 500 nm in width, dimensions commonly used for single-mode silicon waveguides[13]. The Bragg gratings were manufactured by adding lateral corrugations on the wider waveguides along a length of 250 μm. Adding a discontinuity in the corrugation at the centre of the corrugated section transformed the Bragg gratings into π-shifted Bragg grating etalons.

After manufacture, the wafer was diced into chips measuring 6 mm × 3 mm × 0.8 mm, with the π-shifted Bragg grating etalons located parallel to the long dimension of the chip at a distance of 250 μm from the long edge (Fig. 1c, 'SWEDs'). The chip was cut perpendicular to the orientation of the π-shifted Bragg grating etalons at a distance of approximately 500 μm from the centre of the discontinuity. The chip facet along the cut was then precision-polished with progressively finer diamond-grit polishing films (grit size from 0.1 μm to 30 μm), followed by a final polish with SiO_2 lapping film (grit size of 0.02 μm). The spectral responses of the π-shifted Bragg grating etalons were monitored during the polishing process (see Methods section 'Process control with spectral response'). The polishing stopped once the spacer length (Fig. 1c, 'spacer') reached the desired value.

After polishing, the chip measured 3 mm × 3 mm × 0.8 mm and the polished facet was coated using a method that uses Ag diamine solution, also known as Tollen's reagent, and dextrose that reduces Ag ions to elemental Ag nanoparticles. Before applying the electroless wet chemical deposition solution, the polished facet was cleaned by sonication in distilled water, isopropanol and subsequently acetone. The facet was immersed in Ag diamine solution and the reaction was activated by adding dextrose solution with 1:1 ratio to Ag diamine solution, resulting in an Ag thickness of about 200 nm (ref. [42]). The changes in reflection spectra of the SWED were monitored continuously during the chemical deposition process. The chip was removed from the chemical solution after obtaining a resonance within the reflection band, and the coated chip was cleaned by rinsing with an adequate amount of distilled water to remove the residue from the chemical reaction. As the thickness of the Ag layer is much smaller than the acoustic wavelengths detected by the SWEDs, it does not induce impedance mismatch between the silicon chip and water interface.

Next, the SWEDs were connectorized to an array of eight single-mode polarization-maintaining fibres (Meisu Technology) designed to excite the transverse electric mode in the SWEDs. Each SWED was connected to a single on-chip focusing grating coupler[43] (Fig. 1b, 'interface') via a 15-μm-long adiabatic taper (Fig. 1d, 'taper') and a 450-nm-wide rectangular Si waveguide (Fig. 1d, 'waveguides'). The connectorization of the chips was performed in-house by aligning the fibre array and gluing it with epoxy over the focusing grating couplers; the spacing of the grating couplers and the pitch of the fibre array were both 127 μm. The fibres in the array act as input–output ports for the sequential interrogation of the SWEDs and were individually connected to an interrogation scheme based on a tuneable continuous-wave laser in the C-band (see Fig. 1a).

Process control with spectral response

The polishing process had to be carefully monitored to ensure that the optical cavity was not polished away and to precisely control the spacer length. It is possible to correlate the shape of the reflection spectra to the length of the spacer. Extended Data Fig. 1a shows the reflection spectra of a π-shifted Bragg etalon with $\Delta w = 40$ nm, with a bandgap roughly 5 nm wide (at FWHM) and an approximately 94-pm-wide resonance (at FWHM) in the middle of the bandgap. When one of the Bragg gratings is almost entirely polished away, the confinement efficiency of the light in the cavity is drastically reduced and the resonance vanishes while the spectrum still maintains some degree of asymmetry due to the discontinuity (Extended Data Fig. 1b). When the facet of the chip is coated with a thin reflective film (roughly 200 nm of Ag), the optical confinement efficiency is restored and a resonance with a similar *Q*-factor to the original appears (Extended Data Fig. 1c).

Device characterization

During SWED characterization and optoacoustic imaging, ultrasonic signals were detected by monitoring variations in the reflected intensity from the SWED. This was done using a continuous-wave laser (C-band, 20 mW; INTUN TL1550-B, Thorlabs) and a high-bandwidth photodiode (detection bandwidth, 1.6 GHz; PDB480C, Thorlabs). Ultrasonic signals were excited using a 532-nm pulsed laser with a maximum pulse repetition rate of 1.2 KHz and pulse width of 1.2 ns (Flare PQ HP GR 2k-500, Innolight). The laser power was attenuated on-demand with neutral density filters inserted along the optical path. A photodiode (DET36A, Thorlabs) located near the laser output triggered signal acquisition by a high-speed 3-GS data acquisition card (GaGe). SWED characterization was performed using the setup in Extended Data Fig. 2a. The excitation laser beam was resized in a telescope and spatially filtered by a 25-μm pinhole, then focused with a microscope objective (PLN 10×, NA = 0.25; Olympus) with an optical focus of approximately 2.2 μm laterally. The SWEDs were characterized with an ultrasonic source and acoustically coupled with few drops of water. The ultrasonic source was generated by means of the optoacoustic effect when the focal spot was aligned to coincide with an optical absorber. The chip was positioned on top of the ultrasonic point source using 3D linear translation stages (MTS50-Z8, Thorlabs; not shown in Extended Data Fig. 2).

The sensitivity was determined using 125-μm-thick vinyl black tape (type 764, 3M) as an optical absorber, illuminated by an average optical power of 0.64 mW. The tape was glued to a microscope coverslip of 150 μm thickness; a 0.5-mm needle hydrophone (Precision Acoustics) was used to calibrate the acoustic source. No signal averaging was done during the calibration.

To characterize the bandwidth, a thin gold film (200 nm) was sputtered onto the microscope coverslip to serve as an optical absorber and ultrabroad acoustic frequency source[44]. The gold film was illuminated by an average optical power of 0.1 mW. SWEDs' responses to the point source were passed through a bandpass filter of [2, 500] MHz and averaged 1×10^3 times to compensate for the extremely weak signals generated from the thin layer.

The spatial response of the SWED was characterized with the same ultrasonic point source using linear translation stages. The SWED was scanned along 4 mm (10-μm step size), and the signals recorded at each position were averaged 300 times with a bandpass filter of [5, 150] MHz. The acceptance angle of the SWED was calculated from the longitudinal wave profile of the spatial response (profile (i) in Fig. 2f) by determining the maximum angle at which the longitudinal signals with intensities of at least −6 dB of the most intense longitudinal signal could be detected.

Optoacoustic imaging

Reflection-mode optoacoustic imaging with the SWEDs was performed using the setup in Extended Data Fig. 2b. The beam from the excitation laser was focused with a fibre collimator (F810SMA-543, NA = 0.26,

$f = 34.74$ mm, Thorlabs) and guided through a multimode optical fibre (M92L02, NA = 0.22, core diameter of 200 µm, Thorlabs). The phantoms were raster-scanned using a set of linear translation stages (MLS203-1, Thorlabs; not shown in Extended Data Fig. 2). The chip was positioned perpendicular to the sample plane with the long dimension of the chip facet parallel to the y axis. The sample plane was illuminated by the beam diverging from the fibre with average optical power of 3.2 mW, forming a spot of around about 1 mm in diameter below the chip.

The first sample was a black polystyrene suture with a dimeter of 30 µm (Dafilon Polyamide, B. Braun Melsungen), the second was a 1951 USAF resolution test target (Edmond Optics) and the third was a slim-bar gold fine hexagonal mesh (G400HH, SPI Supplies). The samples were acoustically coupled to the chip with water. The raw data acquired was averaged 50 times and filtered using a bandpass filter of [2, 300] MHz when imaging with SWED$3_{40\mathrm{nm}}$ and [20, 350] MHz when imaging with SWED4.

The first sample was imaged with a step size of 1 µm (Fig. 3b, d), the second sample was imaged with step sizes of 2 µm (Fig. 3f, h), or 300 nm in the enlarged section (Fig. 3j), and the third sample was imaged with step sizes of 5 µm (Fig. 3h) and 2 µm (Fig. 3i). The edge-spread function (ESF) of SWED4 was measured with a step size of 200 nm, and that of the confocal microscope was measured with a pixel size of 50 nm.

The acoustic fringe pattern imaging (Fig. 4) was performed using the setup in Extended Data Fig. 2a and SWED$3_{40\mathrm{nm}}$. The pattern was generated by defocusing the objective to form a spot approximately 260 µm in diameter on the gold-film sample used for bandwidth characterization, illuminated with the same pulsed laser. The SWED was linearly scanned over 400 µm with a step size of 1 µm along the diameter of the spot. The samples were acoustically coupled to the chip with water and the raw data acquired was averaged 300 times and filtered using a bandpass filter of [2, 500] MHz.

Image reconstruction

During imaging, the SWED is raster-scanned over a sample. For each position, depth-resolved, time-sampled optoacoustic signals (A-scan) are acquired over a volume defined by the acceptance angle of the detector. Therefore, at each time point, the detector measures an integration of responses from a spherical shell rather than from a single point in space. To obtain a diffraction-limited image, a backward problem has to be solved, otherwise the image will be heavily blurred and distorted by signals originating not in the direct line of sight of the detector. We used a back-projection algorithm in the frequency domain to solve the backward problem and reconstruct the diffraction-limited image[45]. All the far-field images were reconstructed except the interference patterns in Fig. 4.

Long-term detection stability

During long imaging sessions, the temperature stability of the optical resonance is of concern as it can affect the sensitivity of the detector. For this reason, we monitored the long-term detection stability of the SWED, over 15 min, by examining the ultrasonic signals emitted from a black vinyl tape under the same experimental parameters as for the imaging experiments. This is a common approach for characterizing sensor stability[23]. The results are shown in Extended Data Fig. 3. The standard deviation of the values observed, without application of digital filtering (Extended Data Fig. 3, blue curve), was only 3.9% of the root mean square (r.m.s.). During imaging, a band-pass filter is applied (Extended Data Fig. 3, red curve), which reduces this number to only 0.3% of the r.m.s. This latter value corresponds to a resonance shift of only about 1.5 pm. This finding confirms that the SWEDs are stable during long imaging sessions, without the use of additional stabilization, since a typical SWED resonance, as reported here, is roughly 70–150 pm in FWHM.

Optical point-spread function

We used an Olympus IX-83 confocal microscope with a UPLSAPO60XW NA = 1.2 water immersion objective (Olympus). The point-spread function (PSF) of the confocal microscope (Fig. 3p) was obtained by raster-scanning and collecting the reflected light from a single 150-nm gold nanoparticle (part no. 746649, Sigma Aldrich) immobilized on a 170-µm coverglass (no. 1.5H, Paul Marienfeld) using a layer of poly-L-lysine (part no. P8920-100ML, Sigma Aldrich). The nanoparticle was illuminated by a pulsed laser with a wavelength of 485 nm (LDH-D-C-485, PicoQuant) driven by a corresponding driver (PDL 800-D, PicoQuant), collimated and then coupled to the side port of the confocal microscope. The reflected light was separated from the illumination light using a 2-µm thin pellicle beam splitter (BP108, Thorlabs) and redirected towards a photomultiplier-tube module (H10723-01, Hamamatsu Photonics). The voltage signal from the photomultiplier tube was digitalized by a data acquisition card (PCIe-6353, National Instruments); the raster-scanning was performed with scanning mirrors (dynAXIS 421, ScanLab).

Acoustic PSF

To estimate the acoustic PSF of the SWED, we performed the following procedure. The optical ESF and optical PSF were acquired with the confocal microscope, with pixel sizes of 50 nm and 10 nm, respectively; the acoustic ESF was acquired with SWED4 in the near field with a pixel size of 200 nm. The optical ESF and the acoustic ESF were interpolated to match the sampling rate of the optical PSF; the optical PSF was then deconvoluted from the optical ESF using the Richardson–Lucy algorithm with 10 iterations to obtain the true edge profile of the target sample. The profile and the acoustic ESF were normalized and their slopes were linearly fitted with $R^2 > 0.92$; the profile was then deconvoluted from the acoustic ESF using the Richardson–Lucy algorithm with 10 iterations to obtain an estimate of the acoustic PSF (Fig. 3p).

NEP

The needle hydrophone used for the acoustic source calibration has an approximately flat spectral response in the frequency band [5, 30] MHz. Because calibrated ultrasound transducers in higher frequencies and wider frequency bands are not available, the NEP value of the SWED has to be extrapolated.

We divide the large bandwidth of the SWED into smaller sections with widths identical to the calibrated frequency band of the hydrophone (25 MHz). The NEP in each frequency band is $\mathrm{NEP}_i = N_i/S_iP_i$, where i is the index indicating the frequency band, N_i is the noise in band i, S_i is the detected signal in band i and P_i is the acoustic pressure in band i. If band i overlaps with the frequency response of a calibrated transducer, NEP_i can be measured. To calculate the NEP of the SWED in a different frequency band j, we make two assumptions: $N_i \approx N_j$ and $P_i \approx P_j$. Then, NEP_j can be extracted from:

$$\mathrm{NEP}_i/\mathrm{NEP}_j \approx S_i/S_j \qquad (2)$$

These assumptions hold particularly for the ultrasonic point source used for bandwidth characterization[44]. By applying a band-pass filter to the signal recorded by SWED1 in Fig. 2b we get: $\mathrm{NEP}_{[5,30]\mathrm{MHz}} = 2.1\mathrm{NEP}_{[24.5,49.5]\mathrm{MHz}}$. This result means that if two sources were to emit the same amount of pressure in the [5, 30]-MHz and the [25.5, 49.5]-MHz bands, the SWED response would be stronger by a factor of 2.1 in the latter case.

To determine $\mathrm{NEP}_{[5,30]\mathrm{MHz}}$, first SWED1 and then the needle hydrophone were exposed to the same ultrasonic source; the recorded signals without averaging are depicted in Extended Data Fig. 4. Using the hydrophone, we found that the source was emitting 8.4 KPa in the [5, 30]-MHz frequency band. Dividing this value by the pressure amplitude recorded by SWED1 in this band reveals a sensitivity of 80.7 mv KPa^{-1} that corresponds to $\mathrm{NEP}_{[5,30]\mathrm{MHz}} = 94$ Pa. From equation (2), we find that $\mathrm{NEP}_{[24.5,49.5]\mathrm{MHz}} = 45$ Pa (9 mPa $\mathrm{Hz}^{-1/2}$). These values are 5.4 and 1.3 times larger than the r.m.s. value of the induced pressure on the SWED aperture due to thermal noise in water, in the respective measurement bands[46]. The noise of the SWED was evaluated before the arrival of the ultrasonic signals (less than 0.3 μs) or long after (more than 2 μs). This excludes the oscillations originating from ultrasound reflections (Extended Data Fig. 4, 'reflections') between the SWED and the coverslip as a result of acoustic impedance mismatch.

Data availability

The data that support the findings of this study are available from the corresponding authors on reasonable request.

Code availability

The code used to analyse the data is available from the corresponding authors on reasonable request.

42. Shacham-Diamand, Y., Inberg, A., Sverdlov, Y. & Croitoru, N. Electroless silver and silver with tungsten thin films for microelectronics and microelectromechanical system applications. *J. Electrochem. Soc.* **147**, 3345–3349 (2000).
43. Van Laere, F. et al. Compact focusing grating couplers between optical fibers and silicon-on-insulator photonic wire waveguides. In *Proc. Conference on Optical Fiber Communication and the National Fiber Optic Engineers Conference (OFC/NFOEC)* (ed. Li, C.) 1–3 (IEEE, 2007).
44. Strohm, E. M., Berndl, E. S. L. & Kolios, M. C. High frequency label-free photoacoustic microscopy of single cells. *Photoacoustics* **1**, 49–53 (2013).
45. Köstli, P. K., Martin, F., Hans, B. & Heinz, W. P. Temporal backward projection of optoacoustic pressure transients using Fourier transform methods. *Phys. Med. Biol.* **46**, 1863–1872 (2001).
46. Mellen, R. H. The thermal-noise limit in the detection of underwater acoustic signals. *J. Acoust. Soc. Am.* **24**, 478–480 (1952).

Acknowledgements We thank D. Razansky, H. Estrada, R. J. Wilson, C. Zakian and A. Rosenthal for comments and discussions. We thank the staff at IMEC (Leuven, Belgium), ePIXfab (Ghent, Belgium) and Ara Coatings (Erlangen, Germany) for technical expertise at various steps during the manufacturing process. We acknowledge financial support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement numbers 694968 (PREMSOT), 667933 (MIB) and 732720 (ESOTRAC), and from the Deutsche Forschungsgemeinschaft (DFG; Gottfried Wilhelm Leibniz Prize 2013; NT 3/10-1).

Author contributions R.S. conceived the SWED and its principle of operation, and designed the chip layout. R.S. and G.W. constructed the first SWED prototype. R.S. and O.Ü. performed experimental work leading to the demonstration of imaging performance and detector resolution. O.Ü. polished, coated and connectorized the chips. R.S. and V.N. designed experiments, and R.S. and O.Ü. conducted the imaging experiments. R.S. processed and analysed the data. Q.M. developed the image reconstruction algorithm and contributed to image processing. A.C. advised on the experimental setup and constructed the optical confocal microscope. R.S. and V.N. wrote the manuscript. V.N. supervised the research. All authors read and edited the manuscript.

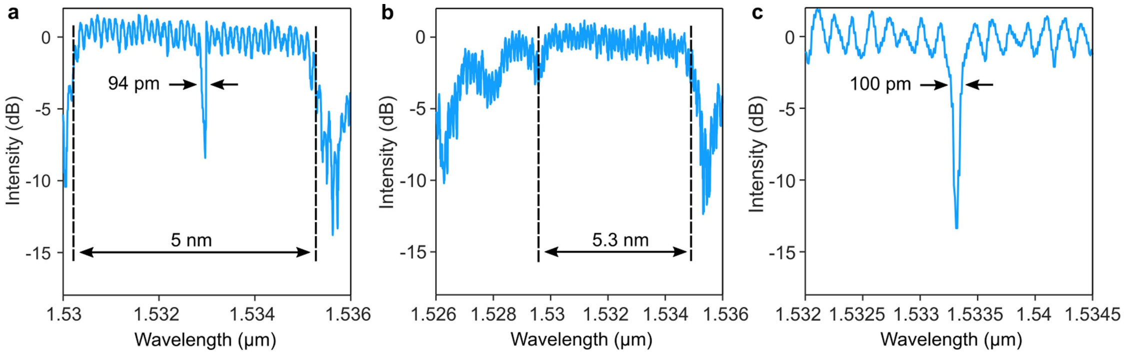
Competing interests V.N. has financial interests in iThera Medical GmbH, Surgvision BV/Bracco S.p.A, I3 Inc. and Spear UG.

Additional information

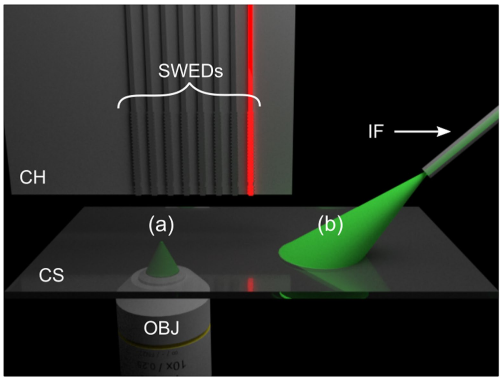
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2685-y.

Correspondence and requests for materials should be addressed to R.S. or V.N.

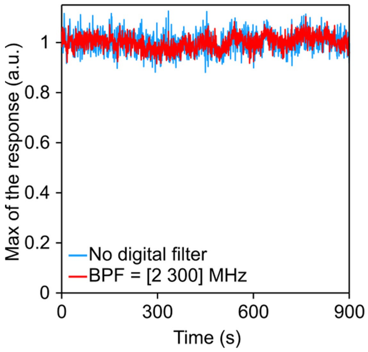
Peer review information *Nature* thanks Ivan Pelivanov and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



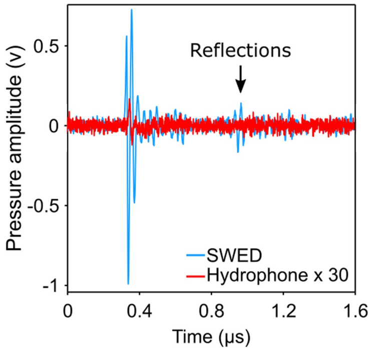
Extended Data Fig. 1 | Process control with spectral response. **a**–**c**, Monitoring the reflection spectrum of a SWED with $\Delta w = 40$ nm (SWED3_{40nm}) during the polishing process of the SOI chip: before the start of polishing (**a**); at the end of polishing and before the application of the Ag coating (**b**); and after the application of the Ag coating (**c**).



Extended Data Fig. 2 | Schematics of the experimental setups. **a**, Characterization setup. An inverted microscope is coupled to a laser source for optoacoustic excitation; the SOI chip (CH) is mounted in a trans-illumination geometry and is raster-scanned over the sample placed on the coverslip (CS) (stages not shown). **b**, Imaging setup. The laser source for optoacoustic excitation is coupled into an optical fibre, which illuminates the sample; the chip is mounted in a reflection-mode illumination geometry. The coverslip holding the sample is raster-scanned while the chip and the illumination fibre (IF) are stationary (stages not shown). In both setups, the SWED interrogation is performed by a tuneable continuous-wave laser. OBJ, microscope objective.



Extended Data Fig. 3 | Long-term detection stability of the SWED. The variation in the maximum values of the detected signal amplitude is shown over 15 min. BPF, band-pass filter; a.u., arbitrary units.



Extended Data Fig. 4 | Hydrophone and SWED responses to an acoustic point source over the frequency band [2, 500] MHz. The hydrophone response has been scaled up by a factor of 30 for visibility. Signals attributed to reflections between the SWED and the sample holder are indicated.

Extended Data Table 1 | The measured optical and acoustic properties of the manufactured SWEDs

	Optical properties					Acoustic properties	
	Spacer length (μm)	Δw (nm)	Resonance wavelength (nm)	FWHM (pm)	Q-factor	f_c (MHz)	Bandwidth -3 dB / -6 dB (MHz)
SWED1	26	40	1534.09	69	2.2×10^4	37	103 / 131
SWED2	14	40	1533.18	78	1.97×10^4	61	107 / 172
SWED3-40	9	40	1532.47	100	1.5×10^4	88	166 / 230
SWED3-30	9	30	1528.8	150	1.0×10^4	-	- / -
SWED4	3.5	40	1531.59	87	1.76×10^4	121	100 / 195

Article

Plasmonic enhancement of stability and brightness in organic light-emitting devices

https://doi.org/10.1038/s41586-020-2684-z

Received: 13 February 2020

Accepted: 2 July 2020

Michael A. Fusella[1], **Renata Saramak**[1], **Rezlind Bushati**[1], **Vinod M. Menon**[1], **Michael S. Weaver**[1], **Nicholas J. Thompson**[1]✉ & **Julia J. Brown**[1]

The field of plasmonics, which studies the resonant interactions of electromagnetic waves and free electrons in solid-state materials[1], has yet to be put to large-scale commercial application[2] owing to the large amount of loss that usually occurs in plasmonic materials[3]. Organic light-emitting devices (OLEDs)[4–7] have been incorporated into billions of commercial products because of their good colour saturation, versatile form factor[8] and low power consumption[9], but could still be improved in terms of efficiency and stability. Although OLEDs incorporating organic phosphors achieve an internal charge-to-light conversion of unity[10], their refractive index contrast reduces the observable fraction of photons outside the device to around 25 per cent[11–13]. Further, during OLED operation, a localized buildup of slow-decaying[14] triplet excitons and charges[15] gradually reduces the brightness of the device in a process called ageing[16,17], which can result in 'burn-in' effects on the display. Simultaneously improving device efficiency and stability is of paramount importance for OLED technology. Here we demonstrate an OLED that uses the decay rate enhancement[18] of a plasmonic system to increase device stability, while maintaining efficiency by incorporating a nanoparticle-based out-coupling scheme to extract energy from the plasmon mode. Using an archetypal phosphorescent emitter, we achieve a two-fold increase in operational stability at the same brightness as a reference conventional device while simultaneously extracting 16 per cent of the energy from the plasmon mode as light. Our approach to increasing OLED stability avoids material-specific designs[19–22] and is applicable to all commercial OLEDs that are currently used for lighting panels, televisions and mobile displays.

Surface plasmons exist at the interface between a metal and the surrounding dielectric environment. These collective oscillations of electrons along the metal surface result in large electric fields and in an orders-of-magnitude improvement in decay rate[18,23] over the visible and near-infrared spectral ranges, making them ideal for use with OLEDs. Typically, the non-radiative surface plasmon mode of a metallic OLED electrode is considered a loss pathway because the quenched exciton energy is dissipated as heat. Much of the work in OLEDs has focused on minimizing this loss[24] by locating the emitter at a large distance from the metallic electrodes or changing the average dipole orientation[11–13]. By contrast, we intentionally couple energy to the surface plasmon mode of the OLED cathode to decrease the excited-state transient and steady-state exciton density of our device. We utilize the phosphorescent emitter, *fac*-tris(2-phenylpyridine) Ir(III) (Ir(ppy)$_3$), hosted by 2,4-diphenyl-6-bis(12-phenylindolo)[2,3-a] carbazole-11-yl)-1,3,5-triazine (DIC-TRZ). We schematically depict the plasmonic OLED in Fig. 1a, where the emissive layer is within 20 nm of the Ag cathode to intentionally couple to the surface plasmon mode to improve the decay rate constant. Light is subsequently out-coupled by randomly arranged Ag nanocubes (Fig. 1b) separated from the Ag cathode by a dielectric layer. We term this device 'plasmon nanopatch antenna' (NPA) even though our nanoparticle out-coupling scheme differs from previous NPA architectures in that our emissive material does not reside in the gap between the metal film and the nanoparticle[25–29]. At accelerated constant-current density ageing at 80 mA cm^{-2}, the plasmon NPA achieves a nearly three-fold stability increase (Fig. 2a) compared to a reference OLED incorporating organic phosphors (PHOLED) ('standard PHOLED') that is similar in layer structure to commercial devices. This stability enhancement occurs despite thinning the emissive layer (EML), which typically decreases the stability of the device[19], from 400 Å in the reference device to 50 Å in the plasmon NPA. This stability enhancement is more pronounced when comparing the plasmon NPA to a device with a 50-Å EML ('thin-EML PHOLED') at the same distance from the metal cathode as the standard PHOLED, whereby the plasmon NPA is four times more stable. We note that, despite the thin device architecture, the plasmon NPA shows no evidence of shorting during the life test (see Supplementary Fig. 7).

This dramatic enhancement in device stability is achieved without loss in efficiency. In Fig. 2b we plot the external quantum efficiency (EQE), which is the number of photons observed in air per injected charge. The standard and thin-EML PHOLEDs have 1,000-Å-thick Al or Ag cathodes, giving bottom emission (BE) through the transparent

---

[1]Universal Display Corporation, Ewing, NJ, USA. ✉e-mail: nthompson@oled.com

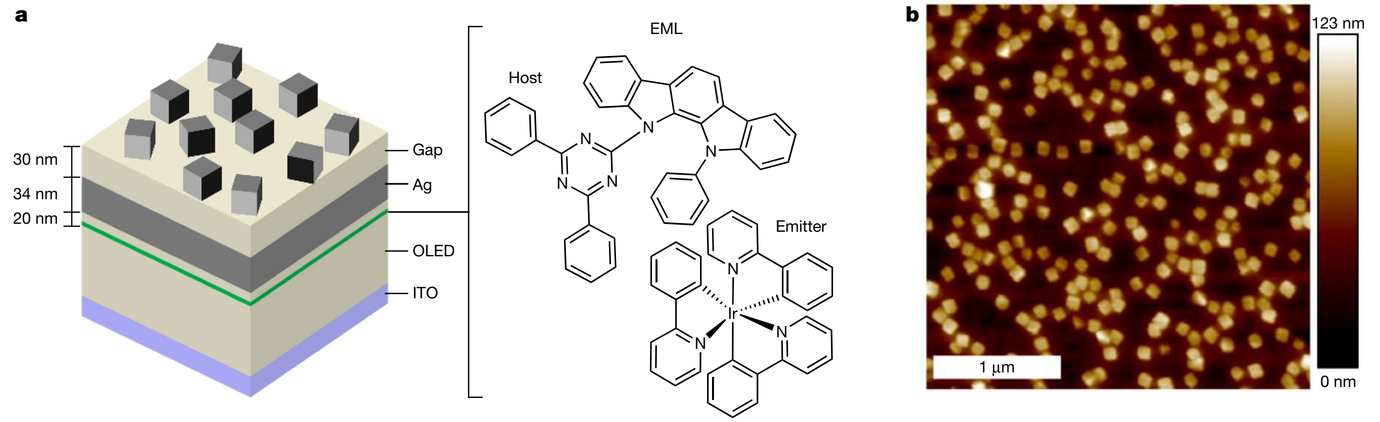


Fig. 1 | Plasmonic device diagram and nanocube morphology. a, Schematic of the plasmon NPA, with relevant layer thicknesses annotated. The EML position and width within the OLED are denoted by the green line. The chemical structures of the EML components, host (DIC-TRZ) and emitter ($Ir(ppy)_3$), are also presented. **b**, Atomic-force micrograph of Ag nanocubes spun on top of the OLED. The fill fraction of Ag cubes is 15%, with a centre-to-centre spacing of ~200 nm. ITO, indium tin oxide.

anode. The EQEs of these devices are about 13% at 10 mA cm^{-2}. Whereas the plasmon NPA uses a transparent anode, it additionally converts energy coupled to the surface plasmon mode of the 340-Å-thick Ag cathode to photons using randomly arranged silver nanocubes, resulting in light emission from the top of the device (top emission, TE). The EQE for the light emitted from the top of the plasmon NPA is 8%, whereas the same device without nanocubes ('plasmon non-NPA') has a TE EQE of only ~1%, highlighting the role of the nanocubes in out-coupling (see Table 1). The choice of a simultaneous top- and bottom-emitting architecture for the plasmon NPA is intentional, because it distinguishes energy that couples into the plasmon mode and is scattered out (TE), from energy that does not couple into the plasmon mode (BE). By measuring both TE and BE light, we can determine the efficiency of energy extraction from the plasmon mode, as discussed later. A commercial application of this device would be designed to eliminate any BE light by coupling all the excitons to the plasmon mode or by using an opaque metal anode to reflect BE light towards the top of the device (see example structures in Supplementary Fig. 17).

To put the results presented in Fig. 2a, b into context, we examine the exciton dynamics inside the EMLs of each of the three devices under investigation. From the curve of normalized EQE versus current density (inset of Fig. 2b) we see that the plasmon NPA maintains its EQE at high current densities better than the reference devices. The reduction in EQE observed at high current density, termed 'roll-off', is attributed to multi-particle interactions, including triplet–triplet annihilation (TTA)[6] and triplet–polaron annihilation (TPA)[15], given that both triplet and charge densities increase at high current density. The thin-EML PHOLED demonstrates greater efficiency roll-off than the standard PHOLED owing to the increased steady-state triplet concentration that arises from confinement of the triplet excitons to the 50-Å EML. By contrast, the plasmon NPA, which has greater coupling to the Ag cathode, shows less efficiency roll-off than both the standard and the thin-EML PHOLEDs, even with a 50-Å EML. We attribute this reduced roll-off to a lower concentration of triplet excitons due to the reduced excited-state lifetime that arises from plasmon coupling. To confirm this assertion, we measure the transient electroluminescence (EL) of the standard, thin-EML and plasmon NPA (Fig. 2c) and find that the plasmon NPA has the shortest decay time. Although the decay curves for all three devices require bi-exponential fitting, they are largely dominated by the fastest decay component, τ_1. The standard PHOLED, thin-EML PHOLED and plasmon NPA have EL transients (τ_1) of 521 ns, 404 ns and 267 ns, respectively (see Supplementary Figs. 9, 10 and Supplementary Table 3 for further fitting details). The thin-EML PHOLED has a reduced τ_1 relative to the standard PHOLED owing to the higher likelihood of TTA and TPA interactions resulting in triplet quenching[6]. The plasmon NPA has a considerably shorter τ_1 owing to coupling of the triplet energy to the surface plasmon mode. With reduced excited-state lifetime, the steady-state population of triplet excitons is lowered,

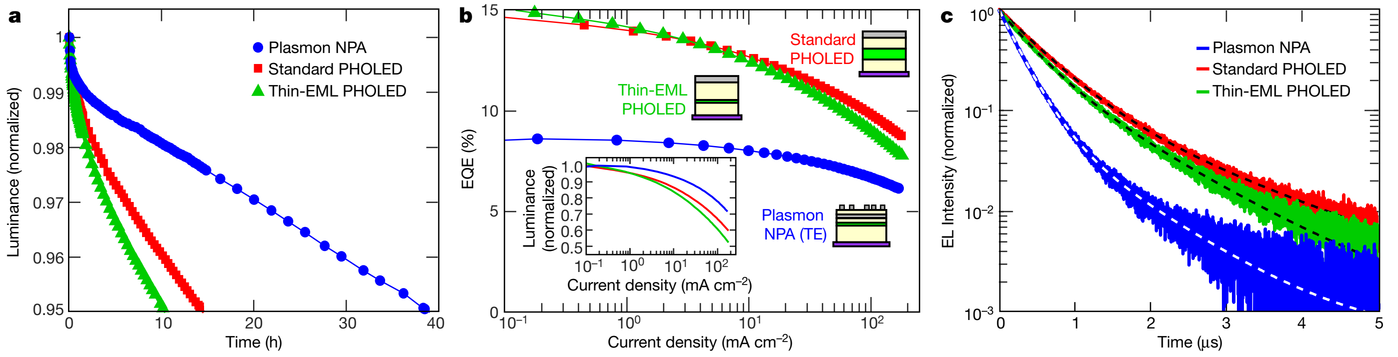


Fig. 2 | Plasmon-enhanced lifetime and efficiency. a, Accelerated ageing stability measurement at a fixed current density of 80 mA cm^{-2} for the plasmon NPA (TE), standard PHOLED (BE) and thin-EML PHOLED (BE). **b**, EQE curves of the plasmon NPA (TE), standard PHOLED (BE) and thin-EML PHOLED (BE). The inset shows the EQE curves normalized at 0.1 mA cm^{-2}, demonstrating reduced efficiency roll-off for the plasmon NPA. Schematic depictions of the device stacks are displayed near each EQE curve and indicate variations in the EML thickness and position relative to the cathode. **c**, Transient EL for the plasmon NPA (TE), standard PHOLED (BE) and thin-EML PHOLED (BE), showing reduced excited-state lifetime for the plasmon NPA. The dashed lines mark the bi-exponential fit for each curve. The plasmon non-NPA transient (omitted for clarity) is nearly identical to that of the plasmon NPA (see Supplementary Fig. 8).

Table 1 | Summary of plasmonic and reference device properties

Device	Emission side	1931 CIE (*x*, *y*)	EQE (%)	EQE$_{PD}$ (%)	LT$_{95}$ at 80 mA cm^{-2} (h)	LT$_{95}$ at 10,000 cd m^{-2} (h)	EL decay, τ_1 (ns)
Standard PHOLED[a]	BE	(0.320, 0.623)	12.7 ± 0.3	—	15	76 ± 2	521
Thin EML PHOLED[a]	BE	(0.331, 0.617)	12.5 ± 0.3	—	10	78 ± 2	404
Plasmon non-NPA	TE	(0.256, 0.647)	0.5 ± 0.1	1.0 ± 0.1	24	0.3 ± 0.1	265
Plasmon non-NPA	BE	(0.294, 0.631)	3.8 ± 0.1	3.8 ± 0.1	—	—	—
Plasmon NPA[a]	TE	(0.335, 0.612)	8.0 ± 0.2	7.4 ± 0.1	39	142 ± 3	267
Plasmon NPA	BE	(0.288, 0.634)	4.8 ± 0.1	4.9 ± 0.1	—	—	—
Plasmon NPA	BE + TE	—	12.8 ± 0.3	12.3 ± 0.1	—	363 ± 7	—

EQEs are measured at 10 mA cm^{-2}. EQE$_{PD}$ is measured with a large-area silicon photodiode[34] to account for the non-Lambertian nature of the top emission. The LT$_{95}$ value at 10,000 cd m^{-2} is calculated using the appropriate acceleration factor (see Supplementary Information Note 2) and the LT$_{95}$ (in hours) from accelerated ageing at 80 mA cm^{-2}. An error value of ±0.1 indicates an error of 0.1 or less; see Supplementary Information Note 2 for more details on error calculations. 1931 CIE, International Commission on Illumination 1931 XYZ colour space.

[a] Devices featured in Fig. 2.

leading to fewer TPA and TTA events and therefore reduced roll-off. Similarly, the reduced triplet density (fewer TPA and TTA events) also leads to greater stability in the plasmon-NPA device[15].

The rate enhancement via plasmon coupling often comes at the expense of reduced photon output and thus a decrease in EQE[30]. However, our plasmon NPA maintains high EQE by using an NPA out-coupling scheme consisting of randomly arranged 75-nm Ag nanocubes separated from the planar Ag cathode by a 300-Å-thick organic gap layer. This architecture is in stark contrast to the typical patch-antenna-based approach, where the emitters are placed between the nanocube and the planar metal film to obtain maximum spontaneous emission rate enhancement[18,28,31]. Here, the rate enhancement arises from the surface plasmon coupling to the planar Ag cathode, whereas the Ag nanocubes perform the out-coupling. This enables broadband rate enhancement without compromising on the device architecture, such as requiring the whole OLED to be placed within the gap. In Fig. 3a, we plot the simulated electric-field intensity at 525 nm for devices with and without the NPA, grafted together. The electric-field intensity in air is markedly higher for the device with the NPA and originates from the corners of the Ag nanocubes, confirming that the NPA is the source of the out-coupled light enhancement.

The plasmon NPA shows no net loss in efficiency, given that its total (that is, TE + BE) EQE is about 13%, matching that of the control devices. To understand the efficiency-limiting factors, we use the BE EQE, TE EQE, transient EL and radiative rate of Ir(ppy)$_3$ in the host material to solve for the fraction of excitons converted to light emitted from the top of the device, as described in Supplementary Information Note 1. We determine that only 50% of the triplet excitons are coupled into the surface plasmon mode in the plasmon NPA. With a TE EQE of 8%, this indicates that ~16% of the excitons coupled into the surface plasmon mode are converted to light emitted from the top of the device. The out-coupling efficiency of the plasmon NPA appears to be limited by the mode overlap between the NPA spectra and emission from Ir(ppy)$_3$. In Fig. 3b, we plot the ratio of the TE EL spectrum to the BE EL spectrum (normalized to 1), which represents the NPA out-coupling spectra and matches the diffuse reflectance of the NPA (Extended Data Fig. 1). We find that this spectrum is not well aligned with the emission of Ir(ppy)$_3$, thereby reducing the out-coupling efficiency of the system by acting as a filter. If the spectral overlap of the NPA with the emission of Ir(ppy)$_3$ were increased to 100% and all the excitons were coupled to the silver cathode, we estimate that the TE EQE would increase to ~20% (details in Supplementary Information Note 1). This energy extraction efficiency may still be limited by several factors, including mode cross-coupling and the distribution of Ag cubes. In devices with a different Ag adhesion layer, we achieve a TE EQE of 11% and a total EQE larger than 16%, which is greater than that of our reference devices; see Extended Data Table 1.

Using finite-difference time-domain modelling we calculate the TE EQE of the device for horizontal, vertical and isotropic dipoles to estimate the ultimate efficiency that can be achieved (Fig. 3c). We find a considerable increase in the predicted TE EQE when the Ag nanocubes are included, achieving a ten-fold increase in EQE for an isotropic emitter such as Ir(ppy)$_3$. This is very close to the nearly eight-fold increase in TE EQE observed experimentally. Whereas the TE EQE is substantially modified by the presence of the Ag nanocubes, the decay

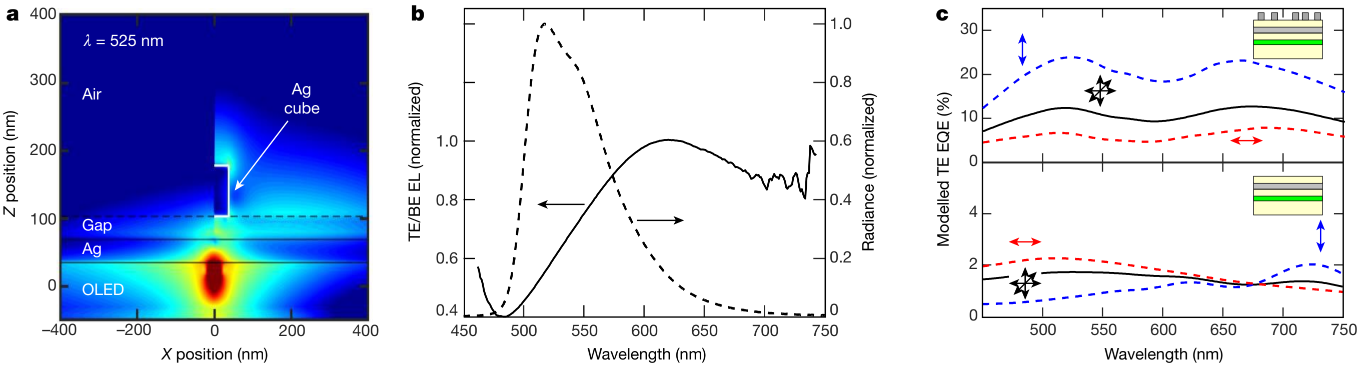


Fig. 3 | Measured and modelled optical properties of plasmon NPA. a, Simulated electric-field intensity maps for a vertical dipole within the OLED without (left) and with (right) a Ag nanocube. Maps are overlaid at 0 nm in the *X* direction. When the Ag cube is present, there is considerable increase in electric-field intensity between the Ag cube and the Ag film, as well as at the corner of the Ag cube, which is the source of radiation to free space. **b**, Plot of the TE/BE EL spectrum (solid line) for the plasmon NPA, showing the spectral shape of the NPA out-coupling. The TE/BE ratio is offset to accentuate that the intrinsic emission spectrum of Ir(ppy)$_3$ (dashed line) is not well aligned with the NPA out-coupling. **c**, Modelled TE EQE versus wavelength for a dipole 20 nm from the Ag cathode with (top) and without (bottom) Ag nanocubes. The dipole orientation—vertical (blue arrows), horizontal (red arrows) or isotropic (black arrows)—is denoted next to each EQE curve. The modelled EQE curves with Ag nanocubes are averages of multiple simulations (see Supplementary Information Note 2 for more details).

rate enhancement is unperturbed, which is consistent with the rate enhancement originating from the Ag cathode only (see Extended Data Fig. 2). For a vertical dipole, the TE EQE is predicted to be greater than 20% over a large wavelength range, where minor variations in EQE originate from the fact that we simulate a single dipole under the cube array (see Extended Data Fig. 3 for more details). Although our modelled TE EQE is sizable, it is still considerably lower than that observed in previous research, which demonstrated plasmon energy extraction efficiencies greater than 60% in gratings[32]. We expect that by going from randomly arranged nanocubes to periodic or quasiperiodic cube arrays, the device out-coupling efficiency could be further enhanced and tailored for directional emission, which would mean that the plasmonic PHOLED can meet, or eventually exceed, standard PHOLED efficiencies.

The display industry typically considers device stability as the number of operating hours required for the eye-perceived brightness to be reduced to 95% of its initial value, LT_{95}. Therefore, the stabilization of a device using the surface plasmon mode is most useful if the extended operational stability can be achieved at a fixed brightness. Using the TE light only, we determine the $LT_{95,TE}$ at 10,000 cd m^{-2} to be 142 h for the plasmon NPA (see Supplementary Information Note 2 for full details), which is nearly double the value of 78 h for both reference devices. Considering that the plasmon NPA TE + BE EQE is equivalent to the control devices, if we assume that all the light could be out-coupled from the top of the device, we find that $LT_{95,TE+BE}$ would be 363 h, which corresponds to more than four times higher stability than the reference devices.

In summary, we demonstrate enhanced OLED stability by improving the decay rate via surface plasmon coupling, a strategy that is usually considered detrimental to the overall device performance. In this first example, we observe an EQE of 8% from the OLED in which energy is intentionally coupled to the surface plasmon of the silver cathode, and demonstrate up to a four-fold improvement in constant-current-density ageing. Importantly, the improved stability of this device architecture can augment material design advances and enable parallel paths of OLED development. We expect fully optimized device geometries to achieve EQEs in excess of 40% and stability enhancements greater[33] than the four-fold improvement that we observed. This enhancement in OLED stability by coupling to plasmons, which has been hitherto considered detrimental, presents a new paradigm for OLED design and paves the way to applications such as low-cost lighting panels, high-luminance applications, ultrafast modulation and the implementation of blue PHOLEDs into displays.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2684-z.

1. Maier, S. A. *Plasmonics: Fundamentals and Applications* (Springer, 2007).
2. Commercializing plasmonics. *Nat. Photon*. **9**, 477–477 (2015).
3. Boltasseva, A. & Atwater, H. A. Low-loss plasmonic metamaterials. *Science* **331**, 290–291 (2011).
4. Tang, C. W. & VanSlyke, S. A. Organic electroluminescent diodes. *Appl. Phys. Lett*. **51**, 913–915 (1987).
5. Burroughes, J. H. et al. Light-emitting diodes based on conjugated polymers. *Nature* **347**, 539–541 (1990); correction **348**, 352 (1990).
6. Baldo, M. A. et al. Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* **395**, 151–154 (1998).
7. Uoyama, H., Goushi, K., Shizu, K., Nomura, H. & Adachi, C. Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* **492**, 234–238 (2012).
8. Forrest, S. R. The path to ubiquitous and low-cost organic electronic appliances on plastic. *Nature* **428**, 911–918 (2004).
9. Pfeiffer, M., Forrest, S. R., Leo, K. & Thompson, M. E. Electrophosphorescent p–i–n organic light-emitting devices for very-high-efficiency flat-panel displays. *Adv. Mater*. **14**, 1633–1636 (2002).
10. Baldo, M. A., Lamansky, S., Burrows, P. E., Thompson, M. E. & Forrest, S. R. Very high-efficiency green organic light-emitting devices based on electrophosphorescence. *Appl. Phys. Lett*. **75**, 4–6 (1999).
11. Kim, S.-Y. et al. Organic light-emitting diodes with 30% external quantum efficiency based on a horizontally oriented emitter. *Adv. Funct. Mater*. **23**, 3896–3900 (2013).
12. Kim, K.-H. et al. Phosphorescent dye-based supramolecules for high-efficiency organic light-emitting diodes. *Nat. Commun*. **5**, 4769 (2014).
13. Shin, H. et al. Sky-blue phosphorescent OLEDs with 34.1% external quantum efficiency using a low refractive index electron transporting layer. *Adv. Mater*. **28**, 4920–4925 (2016).
14. Helfrich, W. & Schneider, W. G. Transients of volume-controlled current and of recombination radiation in anthracene. *J. Chem. Phys*. **44**, 2902–2909 (1966).
15. Giebink, N. C. et al. Intrinsic luminance loss in phosphorescent small-molecule organic light emitting devices due to bimolecular annihilation reactions. *J. Appl. Phys*. **103**, 044509 (2008).
16. Schaer, M., Nüesch, F., Berner, D., Leo, W. & Zuppiroli, L. Water vapor and oxygen degradation mechanisms in organic light emitting diodes. *Adv. Funct. Mater*. **11**, 116–121 (2001).
17. Yamamoto, H. et al. Improved initial drop in operational lifetime of blue phosphorescent organic light emitting device fabricated under ultra high vacuum condition. *Appl. Phys. Lett*. **99**, 033301 (2011).
18. Koenderink, A. F. On the use of Purcell factors for plasmon antennas. *Opt. Lett*., **35**, 4208–4210 (2010).
19. Zhang, Y., Lee, J. & Forrest, S. R. Tenfold increase in the lifetime of blue phosphorescent organic light-emitting diodes. *Nat. Commun*. **5**, 5008 (2014).
20. Tsang, D. P.-K., Matsushima, T. & Adachi, C. Operational stability enhancement in organic light-emitting diodes with ultrathin Liq interlayers. *Sci. Rep*. **6**, 22463 (2016); corrigendum **6**, 26921 (2016).
21. Lee, J. et al. Hot excited state management for long-lived blue phosphorescent organic light-emitting diodes. *Nat. Commun*. **8**, 15566 (2017).
22. Ràfols-Ribé, J. et al. High-performance organic light-emitting diodes comprising ultrastable glass layers. *Sci. Adv*. **4**, eaar8332 (2018).
23. Okamoto, K. et al. Surface-plasmon-enhanced light emitters based on InGaN quantum wells. *Nat. Mater*. **3**, 601–605 (2004).
24. Hobson, P. a., Wedge, S., Wasey, J. e., Sage, I. & Barnes, W. L. Surface plasmon mediated emission from organic light-emitting diodes. *Adv. Mater*. **14**, 1393–1396 (2002).
25. Cesario, J. et al. Coupling localized and extended plasmons to improve the light extraction through metal films. *Opt. Express* **15**, 10533–10539 (2007).
26. Esteban, R., Teperik, T. V. & Greffet, J. J. Optical patch antennas for single photon emission using surface plasmon resonances. *Phys. Rev. Lett*. **104**, 026802 (2010).
27. Belacel, C. et al. Controlling spontaneous emission with plasmonic optical patch antennas. *Nano Lett*. **13**, 1516–1521 (2013).
28. Akselrod, G. M. et al. Probing the mechanisms of large Purcell enhancement in plasmonic nanoantennas. *Nat. Photon*. **8**, 835–840 (2014).
29. Hoang, T. B. et al. Ultrafast spontaneous emission source using plasmonic nanoantennas. *Nat. Commun*. **6**, 7788 (2015).
30. Brütting, W., Frischeisen, J., Schmidt, T. D., Scholz, B. J. & Mayr, C. Device efficiency of organic light-emitting diodes: progress by improved light outcoupling. *Phys. Status Solidi* **210**, 44–65 (2013).
31. Bogdanov, S. I. et al. Ultrabright room-temperature sub-nanosecond emission from single nitrogen-vacancy centers coupled to nanopatch antennas. *Nano Lett*. **18**, 4837–4844 (2018).
32. Worthing, P. T. & Barnes, W. L. Efficient coupling of surface plasmon polaritons to radiation using a bi-grating. *Appl. Phys. Lett*. **79**, 3035–3037 (2001).
33. Ha, D.-G. et al. Dominance of exciton lifetime in the stability of phosphorescent dyes. *Adv. Opt. Mater*. **7**, 1901048 (2019).
34. Im, J. H. et al. Bulk-like Al/Ag bilayer film due to suppression of surface plasmon resonance for high transparent organic light emitting diodes. *Org. Electron*. **33**, 116–120 (2016).

Methods

Spectral measurements (Photo Research Spectrophotometer PR-730) are taken at 10 mA cm^{-2}. Electrical measurements are taken using a source-measure unit (Agilent B2902A). EQE is calculated from the spectrophotometer photon count collected at normal incidence to the sample and assuming a Lambertian angular-emission profile. EQE sweeps are performed using a large-area p–i–n silicon photodiode with a fixed number of voltage source steps and an acquisition delay time at each point to mitigate capacitive charging effects. EQE$_{PD}$ is measured by butt-coupling a large calibrated photodiode to the sample to collect all emitted light passing through the plane of the glass substrate and sweeping the supply voltage. Using the responsivity of the calibrated photodiode, the emission spectrum and the current measured on the photodiode, EQE$_{PD}$ is determined[35].

Measurements of the EL transient are taken using a fibre-coupled photomultiplier tube (Hamamatsu H7826-01) connected to an oscilloscope (Agilent Technologies DSO9104A). The OLED is driven by a square-wave function generator (Rigol DG1022) at 4 kHz with the forward bias set to give a specific current density and a reverse bias of −2 V. The current density is chosen by analysing the curve of normalized EQE versus current density to choose a point between 90% and 95% of the peak EQE, intentionally avoiding any roll-off to prevent any bi-molecular interactions modifying the measured transient time. The 5,000-scan-averaged decay curve is then post-processed to ensure that all data are greater than zero, and subsequently normalized. A bi-exponential fit with baseline is applied to the lower 85% of the data (to avoid fitting anomalies just after the voltage turn-off). The current densities chosen for the standard PHOLED, thin-EML PHOLED and plasmon NPA are 3 mA cm^{-2}, 2 mA cm^{-2} and 6 mA cm^{-2}, respectively.

Patterned ITO-coated glass substrates are cleaned with organic solvents and treated with oxygen plasma and ultraviolet–ozone cleaning. All organic materials and metals are thermally evaporated in vacuum with a base pressure below 10^{-2} Pa. The active area is 2 mm^2. The device structures are as follows (thicknesses in Å). Standard PHOLED: ITO (750)/HIL (100)/HTL (450)/EBL (50)/ DIC-TRZ:Ir(ppy)$_3$(12%) (400)/ ETL (350)/EIL (10)/Al (1,000); thin-EML PHOLED: ITO (750)/HIL (100)/ HTL (450)/EBL (50)/ DIC-TRZ: Ir(ppy)$_3$(12%) (50)/ DIC-TRZ (50)/ETL (650)/EIL (10)/Al (10)/Ag (1,000); plasmon NPA and plasmon non-NPA: ITO (750)/HIL (100)/HTL (450)/EBL (50)/ DIC-TRZ:Ir(ppy)$_3$(12%) (50)/ DIC-TRZ (50)/ETL (100)/EIL (15)/Al (1)/Ag (340)/GAP (300)/Ag nanocubes. The plasmon non-NPA is identical in structure to the plasmon NPA, without the last layer of Ag nanocubes. Here, ITO has a sheet resistance of 16–19 Ω per square, HIL is the hole-injection layer, HTL is the hole-transport layer, EBL is the electron-blocking layer, ETL is the electron-transporting layer, EIL is the electron-injection layer and GAP is the spacer layer between the cathode and Ag nanocubes. Dopings are given in volume per cent. The chemical structures for each layer are given in Supplementary Information Note 3. Device error and reproducibility statistics are available in Supplementary Information Note 2 and Supplementary Table 2. Aluminium was used as an adhesion layer for Ag to improve the quality of the 34-nm-thick film, as has previously been used in OLEDs[34,36].

Commercially sourced 75-nm silver nanocubes (nanoComposix) are concentrated in ethanol solution via centrifugation (5,200*g* for 10 min) to 5 mg ml^{-1} and spin-cast (3,000 rpm, 40 s) atop the gap layer of the OLED. The device is subsequently dried under a rough vacuum for 15 min to remove any remaining solvent, and then encapsulated with a glass/epoxy cap and desiccant in a nitrogen glovebox. Atomic-force microscopy (Bruker Dimension Edge, in tapping mode) is conducted on a glass coated with gap material and Ag nanocubes spun with the same procedure as the device. The diffuse reflectance is measured with an ultraviolet–visible wavelength spectrometer (Shimadzu UV-2600) with an ISR-2600 integrating sphere attachment. The refractive index (*n*) and extinction coefficient (*k*) of our thin Ag film were determined using a Woollam M2000 variable-angle spectroscopic ellipsometer and fitting the data with the Woollam CompleteEASE software to a B-spline function, starting from known Ag *n* and *k* values.

The simulations were carried out with the finite-difference time-domain (FDTD) method using Lumerical FDTD solutions. For the single-cube simulations, a 75-nm cube was placed 30 nm above a 34-nm-thick silver sheet, with a gap layer of refractive index $n = 1.7$—typical for an organic material—between the sheet and the cube. A vertical oscillating electric dipole was placed 20 nm below the silver sheet in a host of refractive index $n = 1.7$ to simulate the emissive layer. For the estimates of TE EQE, a random distribution of 75-nm Ag cubes was placed 30 nm above a 34-nm-thick Ag sheet, with a gap layer of refractive index $n = 1.7$. 900 cubes were used to match the average centre-to-centre spacing of the experiment, which was 200 nm. A single electric dipole oriented either vertically or horizontally was placed 20 nm below the Ag sheet in a host with $n = 1.7$ to simulate the emissive layer. For both simulations, a volume of 6 μm × 6 μm × 1 μm was used with boundary conditions of perfectly matching layers on all sides in order to absorb the light reaching the boundaries and avoid unnecessary reflections. Owing to the complex refractive index of silver, a non-uniform, index-adjusted mesh of 34 mesh cells per wavelength was used throughout the simulation region. Thus, the mesh was most dense in the Ag region and least dense in free space. The rectangular mesh of the system ensured that the cube corners were sharp in order to accurately simulate the plasmonic system interfaces. Frequency-domain field and power monitors were used to calculate the electric-field pattern. A field and power monitor was placed 40 nm above the cubes to capture the top-emission light reaching the far field. This transmitted power was used to calculate the top-emission EQE. The Purcell enhancement was estimated by extracting the power emitted by the dipole in the presence of the plasmonic structure normalized by that in free space.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

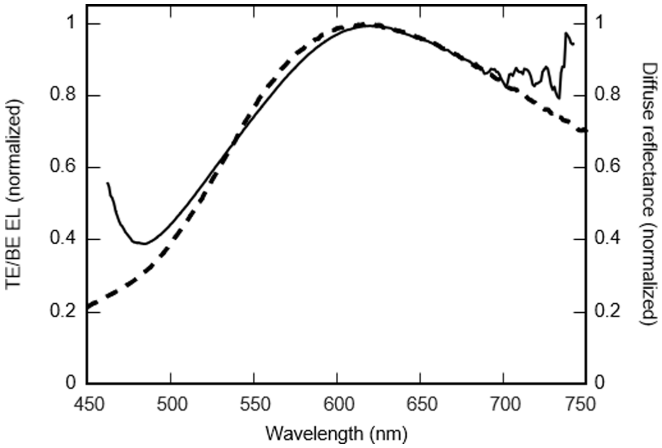
35. Forrest, S. R., Bradley, D. D. C. & Thompson, M. E. Measuring the efficiency of organic light-emitting devices. *Adv. Mater*. **15**, 1043–1048 (2003).
36. Hung, L. S., Tang, C. W., Mason, M. G., Raychaudhuri, P. & Madathil, J. Application of an ultrathin LiF/Al bilayer in organic surface-emitting diodes. *Appl. Phys. Lett*. **78**, 544–546 (2001).

Acknowledgements The authors acknowledge the research staff of Universal Display Corporation (UDC) for discussions and technical assistance.

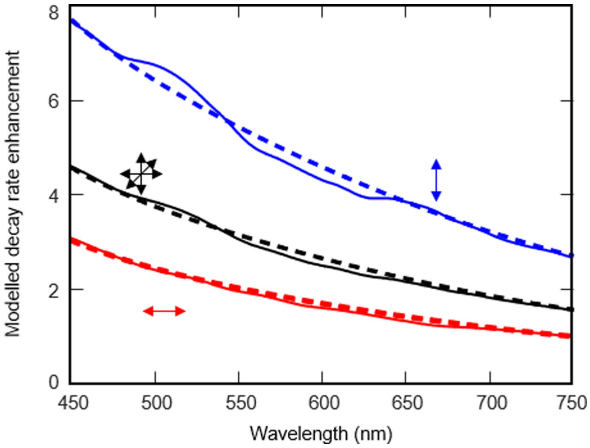
Author contributions M.A.F., M.S.W., V.M.M. and N.J.T. conceived the concepts. M.A.F. and R.S. fabricated the devices and performed the measurements. R.B., M.A.F., V.M.M. and N.J.T. conceptualized the computations and performed the data analysis. M.A.F., R.S., R.B., M.S.W., V.M.M., N.J.T. and J.J.B. contributed to the writing of the manuscript and to discussions of the results.

Competing interests M.A.F., M.S.W., R.S., N.J.T. and J.J.B. are employed at UDC and have personal financial interests via UDC stock ownership and numerous granted and pending patent applications on phosphorescent emitters and OLEDs. Work by R.B. and V.M.M. on this project was completed as part of their consultancy for UDC; they declare no additional competing interests.

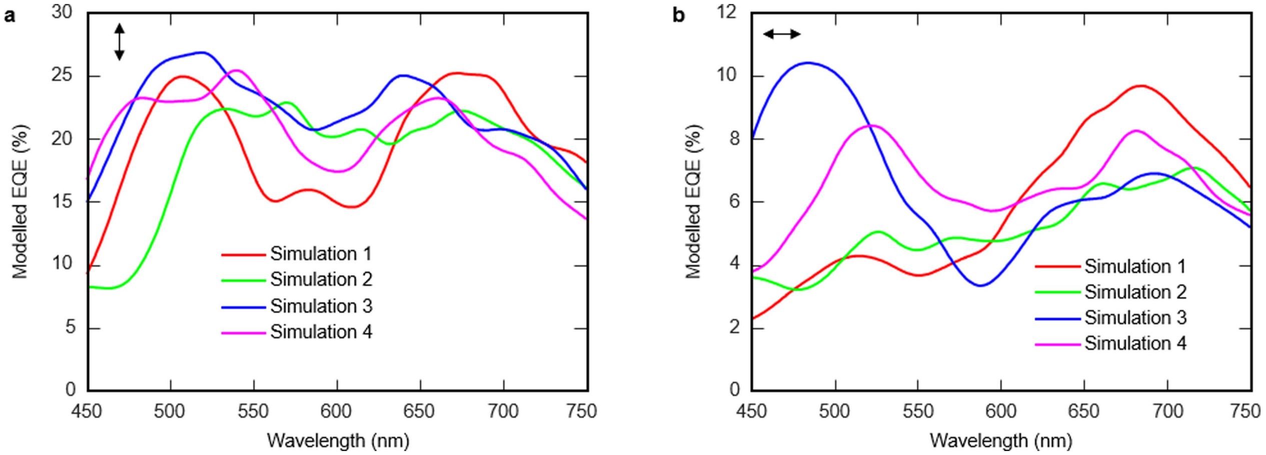
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2684-z.
Correspondence and requests for materials should be addressed to N.J.T.



Extended Data Fig. 1 | Comparison of plasmon NPA resonance on and off a device. Plot of the TE/BE EL spectrum (solid line) for the plasmon NPA, and diffuse reflectance of the plasmon NPA structure itself (dashed line) measured by ultraviolet–visible spectroscopy. There is good agreement between the two measurements, which shows that the TE/BE spectrum presented in Fig. 3b defines the spectral shape of the plasmon NPA out-coupling.



Extended Data Fig. 2 | Decoupling rate enhancement from out-coupling efficiency. Simulated decay rate enhancement of a vertical dipole (blue lines), a horizontal dipole (red lines) and an isotropic dipole (black lines) in the plasmon device structure compared to the corresponding rates in vacuum with (solid) and without (dashed) Ag nanocubes as a function of wavelength. The solid lines are the average of four simulations. Importantly, the rate enhancement is independent of the inclusion of the Ag nanocube out-coupling structure. Therefore, the rate enhancement originates from the Ag cathode, whereas the out-coupling originates from the NPA formed between the Ag cathode, the gap material and the Ag nanocubes.



Extended Data Fig. 3 | Modelled out-coupling efficiencies for several iterations of random nanocube arrays. a, **b**, Modelled TE EQE as a function of wavelength for a vertical (**a**) and a horizontal (**b**) dipole for four simulations with fill factions of 14%. The TE EQE shown in Fig. 3c contains an equal weight of each curve in order to represent the ensemble behaviour of a large distribution of dipole-to-cube-array orientations in the experimental device. The spectral response and maximum efficiency of a single dipole and cube array is highly dependent on the positioning of the cube array relative to the dipole, which is why an average of four curves is presented in Fig. 3c.

Extended Data Table 1 | Summary of device properties for the standard PHOLED, the plasmon NPA and a plasmon NPA (structure #2) incorporating a different adhesion layer for the Ag cathode

Device	Emission Side	1931 CIE (x,y)	EQE (%)	EQE_{PD} (%)	LT_{95} at 80 mA cm^{-2} (h)	LT_{95} at 10,000 cd m^{-2} (h)	EL decay, τ_1 (ns)
Standard PHOLED	BE	(0.320, 0.623)	12.7 ± 0.3	---	15	76 ±2	521
Plasmon NPA	TE	(0.335, 0.612)	8.0 ± 0.2	7.4 ± 0.1	39	142 ± 3	267
Plasmon NPA	BE	(0.288, 0.634)	4.8 ± 0.1	4.9 ± 0.1	---	---	---
Plasmon NPA	BE + TE	---	12.8 ± 0.3	12.3 ± 0.1	---	363 ± 7	---
Plasmon NPA #2	TE	(0.337, 0.613)	10.9 ± 0.2	10.5 ± 0.1	22	148 ± 3	271
Plasmon NPA #2	BE	(0.288, 0.636)	5.5 ± 0.1	6.3 ± 0.1	---	---	---
Plasmon NPA #2	BE + TE	---	16.4 ± 0.3	16.8 ± 0.2	---	334 ±7	---

The EQEs are measured at 10 mA cm^{-2}. EQE_{PD} is measured with a large-area silicon photodiode to account for the non-Lambertian nature of the top emission. The LT_{95} value at 10,000 cd m^{-2} is calculated using the appropriate acceleration factor and the LT_{95} (in hours) from accelerated ageing at 80 mA cm^{-2}. An error value of ±0.1 indicates an error of 0.1 or less.

Reverse and forward engineering of *Drosophila* corneal nanocoatings

https://doi.org/10.1038/s41586-020-2707-9

Received: 20 August 2019

Accepted: 9 July 2020

Mikhail Kryuchkov[1,2], Oleksii Bilousov[2], Jannis Lehmann[3], Manfred Fiebig[3] & Vladimir L. Katanaev[1,2,4] ✉

Insect eyes have an anti-reflective coating, owing to nanostructures on the corneal surface creating a gradient of refractive index between that of air and that of the lens material[1,2]. These nanocoatings have also been shown to provide anti-adhesive functionality[3]. The morphology of corneal nanocoatings are very diverse in arthropods, with nipple-like structures that can be organized into arrays or fused into ridge-like structures[4]. This diversity can be attributed to a reaction–diffusion mechanism[4] and patterning principles developed by Alan Turing[5], which have applications in numerous biological settings[6]. The nanocoatings on insect corneas are one example of such Turing patterns, and the first known example of nanoscale Turing patterns[4]. Here we demonstrate a clear link between the morphology and function of the nanocoatings on *Drosophila* corneas. We find that nanocoatings that consist of individual protrusions have better anti-reflective properties, whereas partially merged structures have better anti-adhesion properties. We use biochemical analysis and genetic modification techniques to reverse engineer the protein Retinin and corneal waxes as the building blocks of the nanostructures. In the context of Turing patterns, these building blocks fulfil the roles of activator and inhibitor, respectively. We then establish low-cost production of Retinin, and mix this synthetic protein with waxes to forward engineer various artificial nanocoatings with insect-like morphology and anti-adhesive or anti-reflective function. Our combined reverse- and forward-engineering approach thus provides a way to economically produce functional nanostructured coatings from biodegradable materials.

Current surface-nanopatterning technologies can produce only limited patterns and use non-eco-friendly methods and materials[7]. By contrast, living organisms have diverse nanocoatings that serve photonic, liquid-handling, bactericidal and other functions[1]. Identifying the mechanisms that govern the formation of these bionanocoatings is important, and a prerequisite for biomimetic applications. Moth-eye nanocoatings provide anti-reflectivity by creating a gradient of refractive indices between those of air and the lens material[1,2]. Corneal nanocoatings are very diverse in arthropods, and can be described by a reaction–diffusion mechanism[4]. This principle of patterning, developed by Alan Turing[5], has numerous applications at the biological macro- and microscales[6]; insect corneas were the first known example of Turing nanopatterns[4].

The Turing mechanism involves interactions between two morphogens, an activator and an inhibitor. We studied insects with fully sequenced genomes—*Drosophila melanogaster*[8] and thirteen other *Drosophila* species[9]—to determine the molecular identity of the morphogens that act as the building blocks of the corneal nanostructures of these insects. We identify the protein Retinin and corneal waxes as the two morphogens. Being intrinsically unstructured, Retinin adopts an induced-fit conformation on wax binding. Retinin and the waxes interact physically and genetically, and they comply with the characteristics of Turing's activator and inhibitor, respectively.

Following this reverse engineering of corneal nanocoatings, we use a forward-engineering approach. We establish low-cost production of Retinin, which after admixture with waxes produces nanocoatings of insect-like morphology on artificial surfaces. Modifying the Retinin and wax components, ratios and surfaces to cover, combined with multilayering, generates highly versatile, stable and eco-friendly surface nanopatterns with diverse properties. Our work identifies how multifunctional nanocoatings are created in nature and translates this knowledge into technological applications. We achieve this through a combination of mathematical simulation, phylogeny, genetics, biochemistry and forward engineering.

Turing nanopatterning in insect corneas

We previously proposed that Turing-based self-assembly governs the formation of insect corneal nanocoatings and their fast evolution rates[2,4,10]. This mechanism describes spatial variation in the concentrations of the slowly diffusing activator and the rapidly diffusing inhibitor[5,6]. Through this mechanism, stable patterns can arise from

---

[1]Department of Cell Physiology and Metabolism, Faculty of Medicine, University of Geneva, Geneva, Switzerland. [2]Department of Pharmacology and Toxicology, University of Lausanne, Lausanne, Switzerland. [3]Department of Materials, ETH Zurich, Zurich, Switzerland. [4]School of Biomedicine, Far Eastern Federal University, Vladivostok, Russia. ✉e-mail: vladimir.katanaev@unige.ch

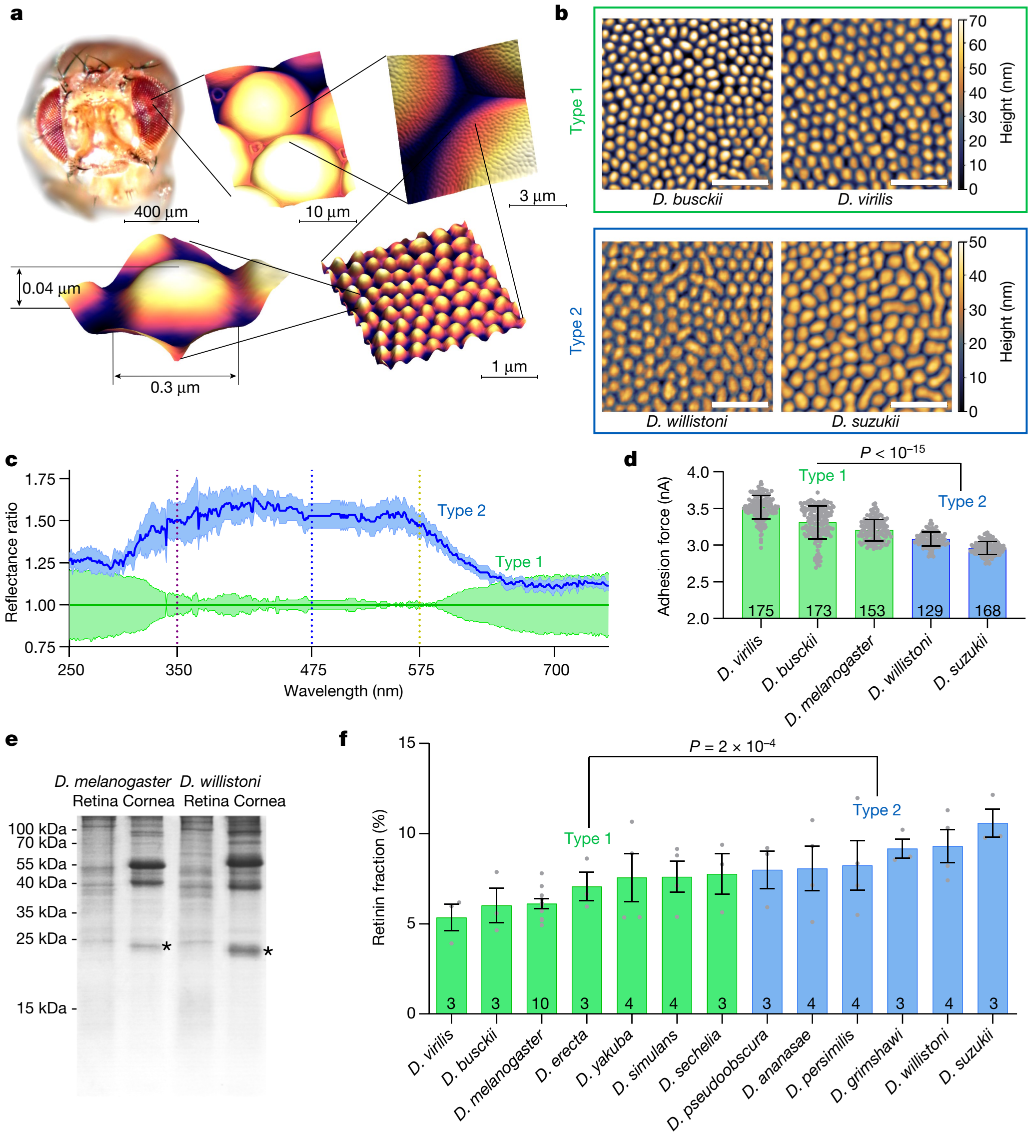


Fig. 1 | Structure, function and composition of corneal nanocoatings across the genus *Drosophila*. **a**, Corneal nanocoatings in *D. melanogaster*. Step-wise increases in magnification are shown, from a macroscale image of a *Drosophila* head to an atomic force microscopy (AFM) image of a single nipple-type nanostructure coating an ommatidial lens. **b**, Two types of corneal nanocoating of the genus *Drosophila*: type 1, individual nipple-like nanostructures seen in *D. melanogaster* and, for example, in *D. busckii* and *D. virilis*; type 2, nipple-to-ridge nanostructures seen, for example, in *D. willistoni* and *D. suzukii*. At least three regions from at least two different animals were analysed using AFM for each species. Scale bars, 1 µm. Note the different height scales. **c**, Ratio of reflection spectra measured for type-2 (*D. sizukii* and *D. willistoni*) and type-1 (*D. busckii* and *D. virilis*) corneas. Three areas for each species were measured; the data for *D. sizukii* and *D. willistoni* were pooled together, as were the data for *D. busckii* and *D. virilis*. Each data point was normalized by the average of the six pooled *D. busckii* and *D. virilis* measurements; the resulting spectra are presented as a reflectance ratio; mean ± s.e.m.; $n = 3$ biologically independent animals; the vertical dotted lines show receptor potential maxima from *D. melanogaster* electroretinograms. **d**, Adhesion-force measurements for different corneas; *n* is indicated on each bar. Individual data are shown as grey circles; the bar and error bars indicate mean ± s.d. The statistical significance of the difference between type-1 (*D. virilis*, *D. busckii* and *D. melanogaster* data pooled together) and type-2 (*D. willistoni* and *D. sizukii* data pooled together) was assessed use a two-tailed *t*-test. **e**, SDS–PAGE of retinal and corneal samples from *D. melanogaster* and *D. willistoni* identifies major corneal-specific protein bands: stars indicate Retinin. The uncropped gel is shown in Supplementary Fig. 11a; similar results were obtained in four independent experiments. **f**, Retinin content in corneas of different *Drosophila* species correlates with the transition from type-1 nanostructures (green) to type-2 nanostructures (blue). Numbers on each bar indicate the number of independent corneal preparations for SDS–PAGE. Individual data are shown as grey circles; the bar and error bars indicate mean ± s.e.m. The statistical significance of the difference in Retinin corneal content between the nipple-only (type 1) and the nipple-to-ridge (type 2) groups was assessed using a two-way ANOVA (no adjustment for multiple comparison).

initial noise-derived concentration inequalities via self-amplification and expansion (Supplementary Fig. 1). We suggest that the activator–inhibitor interaction results in the formation of a reversible complex, which can grow to solid nanoparticles through aggregation and polymerization. We have now developed a 3D model that describes the formation of nanostructures, governed by stochastic secretion of the two morphogens by the corneal surface (Supplementary Note 1, Supplementary Figs. 1–8).

The 3D simulation is based on biological conditions: the insect nanocoatings are generated earlier than other corneal layers; and the future lens forms beneath the nanocoatings[2,11]. Analytical investigation (equations (1)–(3) of Supplementary Note 1) shows the appearance of nanostructures in a thin layer above the source of secretion (Supplementary Fig. 4a). Unequal concentration gradients of the activator and the inhibitor make feasible this thin layer of Turing patterning, as can be illustrated by measuring the growth rates of generic perturbations using the Lyapunov exponent[12] (Supplementary Fig. 4b, c). We further demonstrate the theoretical feasibility of Turing patterns with a scale of tens to hundreds of nanometres (Supplementary Note 1, Supplementary Fig. 6).

The Turing reaction–diffusion mechanism, as detailed below, allowed for predictions that we subsequently validated experimentally, in contrast to, for example, the block-copolymer model (Supplementary Note 1, Supplementary Figs. 5–8). The high level of agreement between the theoretical and experimental findings is not a proof, but an indication that the reaction–diffusion mechanism is behind the formation of bionanocoatings. We used the Turing modelling to guide our experiments, predicting the identities of the molecular components of the nanostructures and their mode of interaction, and applying these mechanisms for forward engineering.

Turing activator across *Drosophila* flies

Species of the genus *Drosophila* (Supplementary Fig. 9a, b) represent fruit flies living in different ecological conditions, including tropical forests, deserts, volcanic islands and human cohabitation[13]. Anti-reflective corneal nanocoatings, first found in moth eyes[14], have subsequently been described in many insect groups[4], including *Drosophila melanogaster*, in which they are built of nipple-type protrusions about 300 nm in width and 40 nm in height[15] (Fig. 1a). By analysing corneal nanocoatings in other *Drosophila* fruit flies, we find that corneal surfaces in the 14 species fall into two principal types. First, nanocoatings composed of individual nipple-like structures ('type-1 nanocoatings') were seen in *D. melanogaster*, and in *D. busckii*, *D. virilis*, *D. sechellia*, *D. erecta*, *D. yakuba* and *D. simulans* (Fig. 1b, Supplementary Note 2, Supplementary Fig. 9c). Second, nanocoatings containing partial fusion of nipple-like structures into ridges ('type-2 nanocoatings') were observed in *D. ananassae*, *D. willistoni*, *D. persimilis*, *D. suzukii*, *D. pseudoobscura* and *D. mojavensis* (Fig. 1b, Supplementary Fig. 9d, e).

Such partial fusion of individual nanostructures into ridges is seen in corneas of other insects as a transitory pattern between nipple-type protrusions and maze-like nanostructures[4]. In agreement with the ease of transition between corneal patterns during insect evolution[4], distinction between type-1 and type-2 nanocoatings in *Drosophila* does not conform with the taxonomical subgrouping within the genus (Supplementary Fig. 9b).

Instead, we find a strong correspondence between the morphological types of nanocoating and their function. The main function of corneal nanostructures is to decrease light reflection from the air–lens interface[2,16]. We observe that type-2 nanocoatings reflect around 50% more light than do type-1 nanocoatings in the ultraviolet and visible parts of the spectrum (Fig. 1c), corresponding to the region of photosensitivity of *Drosophila* photoreceptors[17]. However, type-2 corneal surfaces are significantly less adhesive than type-1 (Fig. 1d). Thus, type-1 nanocoatings (with individual protrusions) have better anti-reflective function but worse anti-adhesive properties than the type-2 nanocoatings (with partially merged structures). The two functions of the nanocoatings—anti-reflectivity and anti-adhesiveness—thus appear to be, to a degree, mutually exclusive. The lifestyle of each *Drosophila* species might determine which function is more relevant for the insect.

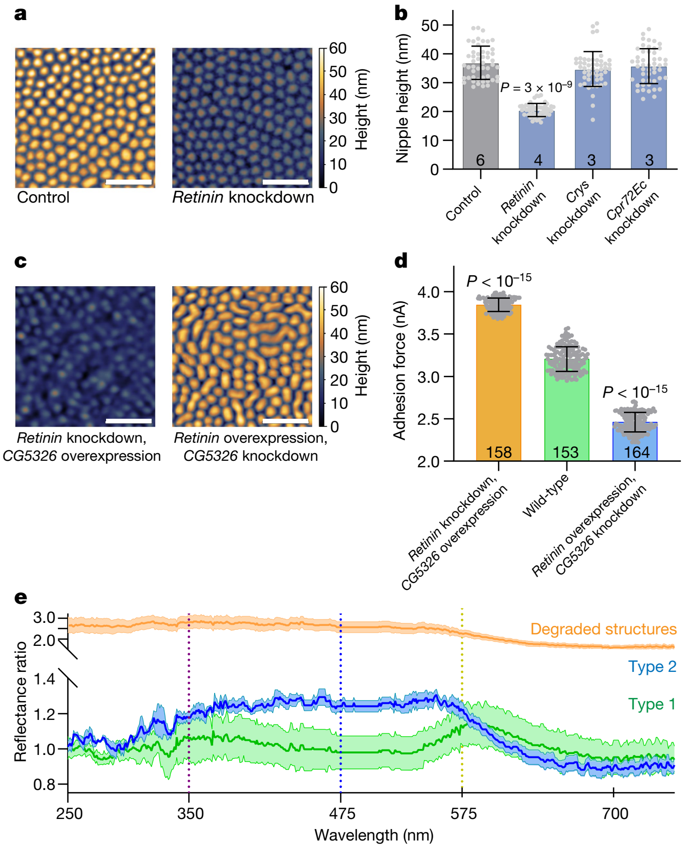


Fig. 2 | Structure and function of the *Drosophila* nanocoatings resulting from genetic manipulations of the Turing activator and inhibitor. **a**, Representative AFM images of control and *Retinin*-knockdown corneas; similar results were obtained in five independent experiments. Scale bars, 1 µm. **b**, Quantification of the height of nanostructures for wild-type, *Retinin*-knockdown, *Crys*-knockdown and *Cpr72Ec*-knockdown genotypes. Individual data are shown as grey circles; the bar and error bars indicate mean ± s.d.; $n = 50$ measurements from three, four or six biologically independent animals (as indicated on each bar). The statistical significance of the difference between control and *Retinin*-knockdown corneas was assessed using a two-tailed *t*-test. **c**, Left, representative corneal nanocoatings from *D. melanogaster* downregulating *Retinin* (knockdown) and overexpressing *CG5326*; degraded, small nipples are seen. Right, overexpression of *Retinin* and downregulation of *CG5326* (knockdown) leads to a clear nipple-to-ridge transition, as in type-2 structures from *Drosophila* species; similar results were obtained in three independent experiments. **d**, Adhesion-force measurement for the degraded corneas (*Retinin* knockdown combined with *CG5326* overexpression), wild-type *D. melanogaster* corneas (type-1 nanostructures) and corneas with merged structures (*Retinin* overexpression combined with *CG5326* knockdown; type-2 nanostructures). Individual data are shown as grey circles; the bar and error bars indicate mean ± s.d. (*n* is indicated on each bar). The statistical significance of the differences between the wild-type and the two mutants was assessed using a two-tailed *t*-test. **e**, Ratio of the reflection spectra measured from corneas of the same three *D. melanogaster* genotypes, normalized by the reflectance of pooled *D. busckii* and *D. virilis* data, as in Fig. 1c; the resulting spectra are presented as a reflectance ratio; mean ± s.e.m.; $n = 3$ biologically independent animals; the vertical dotted lines show receptor potential maxima from *D. melanogaster* electroretinograms.

We suspected that the chemical nature of the Turing activator and inhibitor morphogens behind the corneal nanocoatings could be that of a protein and a lipid, respectively, as these are known to constitute corneal surfaces in different insects[2,18–22]. Following 3D Turing modelling (Supplementary Note 1, Supplementary Fig. 5), we hypothesized that the relative abundance of these morphogens in corneas of different *Drosophila* species determines the type of nanocoating, such that more activator (or less inhibitor) forces nipple fusion into ridges. Detergent treatment of *Bombyx* moth corneas removes nanocoatings and the anti-reflective function they provide, further pointing to the structural role of protein(s) in corneal nanostructures[23]. We find that detergent treatment also removes the nipple-type protrusions from *D. melanogaster* corneas (Supplementary Fig. 10a, b). We therefore aimed to identify the protein component(s) in corneas of *Drosophila* flies.

Analysis of cornea-specific proteins across the *Drosophila* species (Fig. 1e, f, Supplementary Fig. 10c) revealed that one of them (around 25 kDa in size) is significantly more abundant in the species with type-2 than with type-1 nanocoatings. No other corneal protein shows this (or opposite) trend (Supplementary Fig. 10c, d). Using mass spectrometry, we identified this protein as Retinin across the genus *Drosophila* (Supplementary Table 1). In *D. melanogaster*, Retinin has been described to localize exclusively to corneas[24]; it belongs to the insect-specific group of small proteins with an uncharacterized Retinin_C domain (PF04527; http://pfam.xfam.org/family/PF04527).

Thus, Retinin is a likely candidate for the Turing activator. Its increased abundance in several *Drosophila* species correlates (Fig. 1f) with their corneal nanocoatings being of type 2 (nipple-to-ridge fusion pattern). To test this conclusion experimentally and to identify the inhibitor morphogen, we used genetic manipulations in *D. melanogaster*.

Turing inhibitor in *Drosophila* is wax

In addition to Retinin, we identify Crystallin (Crys or Drosocrystallin, about 55 kDa)[25] and Cuticular protein 72Ec (Cpr72Ec, about 40 kDa)[26] as major corneal proteins in *D. melanogaster* (Supplementary Table 2, Fig. 1e, Supplementary Figs. 12, 13a, b), in accordance with a previous study[27]. Consistent with the idea that Retinin is important in the formation of nanocoatings, we found that knockdown of this protein, but not Crys or Cpr72Ec, prominently decreased the size of the nanostructures (Fig. 2a, b, Supplementary Fig. 13a–c). By contrast, overexpression of Retinin induced nipple fusion into corrugated ridges (Supplementary Fig. 13a, d, e), in accordance with the prediction of the Turing model (Supplementary Fig. 5). We found that only a small fraction of overexpressed Retinin could make it into corneas (Supplementary Fig. 13a, d, f, g), which might be the reason for the incomplete merging of the nanostructures (Supplementary Fig. 13e). It also may indicate the existence of additional mechanisms that control Retinin secretion and inclusion into nanocoatings, possibly involving regulation by the Turing inhibitor. We therefore aimed to identify this morphogen.

Lipids in general and waxes in particular have previously been detected in the corneal surfaces of insects[18,22]. We genetically targeted the wax biosynthetic pathway in corneas of *D. melanogaster*. We expected that downregulation of this pathway would produce phenotypes similar to those produced by Retinin overexpression and that upregulation of the pathway would produce phenotypes similar to those of Retinin downregulation (Supplementary Fig. 5). The wax biosynthesis pathway is poorly characterized in arthropods, but has been well studied in mammals[28,29] and plants[30]. We selected two enzymes: fatty-acid elongase, which converts long-chain fatty acids into very-long-chain fatty acids at the beginning of the pathway, and acyltransferase, which makes long-chain ester waxes from long-chain acyl-CoA and long-chain alcohol at the end of the pathway (Supplementary Fig. 14). We genetically targeted the *D. melanogaster* homologues of these enzymes, expressed in the head and/or eye tissue (Supplementary Table 3). For the fatty-acid elongase homologues, knockdown of CG5326, but not of any of the other six candidates, induced fusion of individual nipple-like structures into ridges (Supplementary Fig. 15a), just as Retinin overexpression did. Reciprocally, overexpression of CG5326 induced substantial nipple shrinkage (Supplementary Fig. 15b), just as Retinin knockdown did. Thus, CG5326 is the *Drosophila* homologue of the mammalian fatty-acid elongase and controls wax biosynthesis in corneas.

For the end-point enzyme, similar analysis identified CG1942 as the O-fatty-acyltransferase. Loss-of-function mutation and knockdown of CG1942 (but not another candidate) induced a nipple-to-ridge transformation (Supplementary Fig. 15c), similarly to downregulation of CG5326 or overexpression of Retinin. We therefore conclude that the other component responsible for corneal nanocoating in *Drosophila* is a wax, the synthesis of which is regulated by CG5326 (fatty-acid elongase, the entry-point enzyme) and CG1942 (O-fatty-acyltransferase, the end-point enzyme).

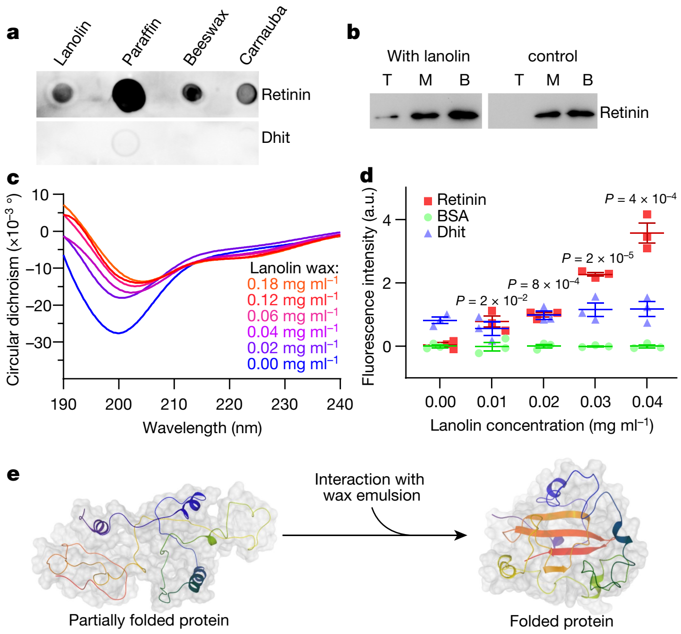


Fig. 3 | Induced folding of Retinin on direct binding to waxes. a, Direct interaction of recombinant Retinin with waxes, shown by a lipid-strip assay. The control protein (Dhit) does not bind waxes; similar results were obtained in two independent experiments. **b**, In the flotation assay, Retinin is found in the top fractions (T) only after addition of lanolin, and in the middle (M) and bottom (B) fractions otherwise. Uncropped membrane and western blots are shown in Supplementary Fig. 11b; similar results were obtained in two independent experiments. **c**, Far-ultraviolet circular-dichroism spectra of Retinin admixtures with different concentrations of the lanolin wax emulsion. Each curve is an average from four readings. **d**, ThT fluorescence at various admixtures of proteins (Retinin, BSA or Dhit) and the lanolin emulsion at different concentrations. Individual data ($n = 3$ biologically independent samples) are shown as coloured symbols; error bars indicate mean ± s.d. The statistical significance of the differences from the control without wax emulsion was assessed using a two-tailed *t*-test. **e**, Illustration of possible changes in the Retinin structure induced by interaction with waxes, based on the Retinin structures predicted by I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER); the colour gradient denotes the direction from the N (violet) to C (red) ends of the polypeptide chain.

The activator and inhibitor of the Turing model interact physically and negatively influence each other. We therefore tested whether Retinin and the components of the wax biosynthetic pathway reveal opposing genetic interactions. We analysed nine genetic interaction conditions, combining overexpression and knockdown of the fatty-acid elongase CG5326 and Retinin (Fig. 2c, Supplementary Fig. 16). Two extremes emerged. The first combined CG5326 upregulation with Retinin knock-down and led to strong degradation of the nanostructuring (Fig. 2c, left). The second combined overexpression of Retinin with knockdown of CG5326 and led to well-developed ridges similar to those of *D. suzukii* corneas (compare Fig. 2c and Fig. 1d). Other combinations in the genetic interaction matrix produced phenotypes between these two extremes, in concordance with the Turing simulations (Supplementary Figs. 5c, e and 16).

We next analysed functional consequences of the two extreme genetic interactions. Similarly to the findings with type-2 versus type-1 nanocoatings of different wild-type *Drosophila* species (Fig. 1e, f), we found that strong nipple-to-ridge merging induced by Retinin overexpression and CG5326 downregulation significantly

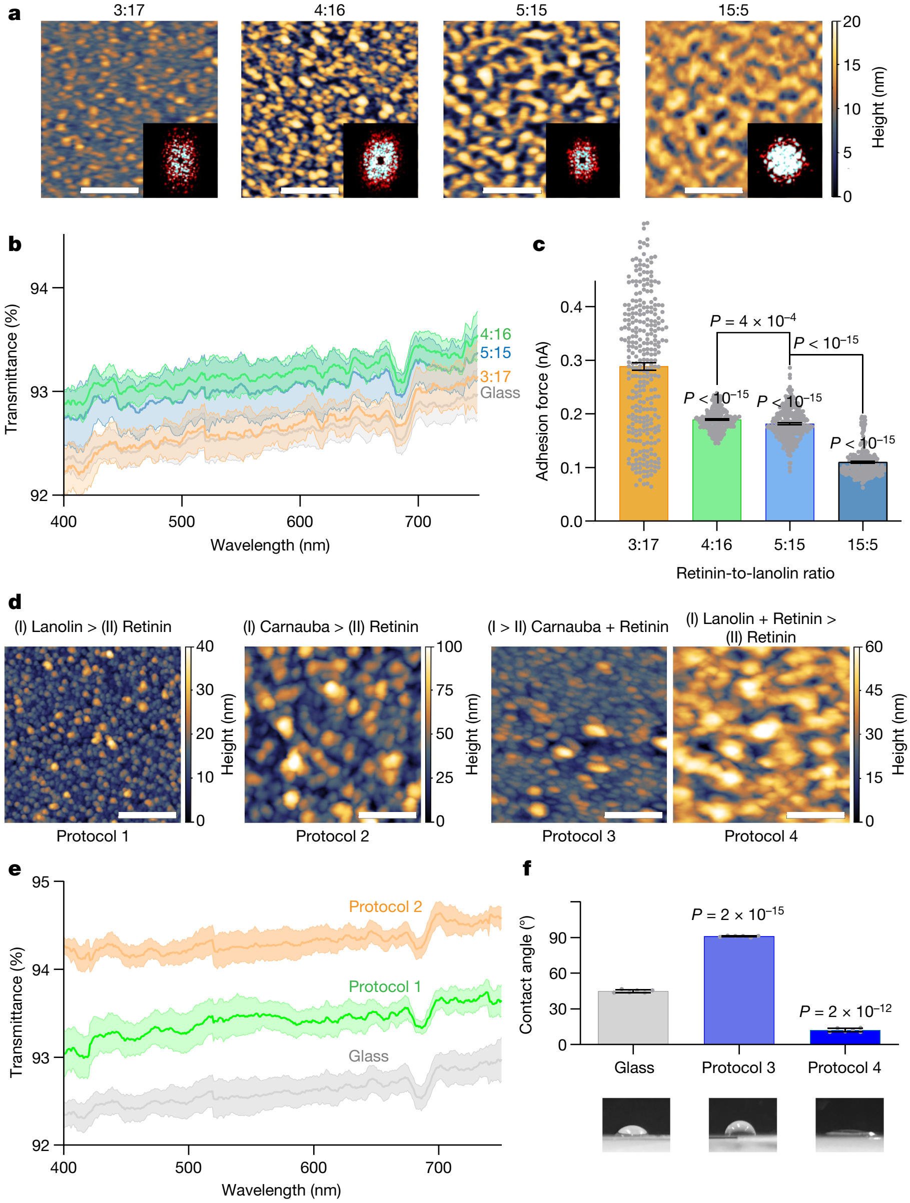


Fig. 4 | In vitro production of insect-like nanocoatings. a, Representative (from two independent experiments) AFM images of nanocoatings formed on glass for different admixtures of Retinin and lanolin (Retinin-to-lanolin ratios are labelled). Two-step coating was applied (Methods; see Supplementary Fig. 17c for single-layer nanocoatings). Insets, Fourier transform spectra of the images. Scale bars, 1 µm. **b**, Transmission spectra for artificial nanocoatings with nipple-like (Retinin-to-lanolin ratio of 4:16) and maze-type (5:15 ratio) structures from **a** reveal improvements compared to glass (grey trace); the 3:17-ratio coating with degraded structures was indistinguishable from glass, as was the dimpled pattern (15:5 ratio, trace not shown). Data are mean ± s.d., relative to transmission through air (100%); *n* = 3 independent experiments. **c**, Adhesion-force measurement for the nanocoatings from **a**. Individual data (*n* = 300 measurements) are shown as grey circles; the bar and error bars indicate mean ± s.e.m. The statistical significance of differences from the 3:17-ratio coating, and between functional nanocoatings, was assessed using a two-tailed *t*-test. **d**, Representative (from two independent experiments) AFM images of nanostructures produced by different protocols (see Methods). The protocols are illustrated above each image, with the materials used and coating layers (by Roman numerals; '>' indicates the sequence of layering) indicated. These coatings provide antireflective (protocols 1 and 2), hydrophobic (protocol 3) or hydrophilic (protocol 4) functionality. Scale bars, 1 µm. **e**, Transmission spectra measured for samples from **d**. Data are mean ± s.d., relative to transmission through air (100%; transmission through uncoated glass is about 92%); *n* = 3 independent experiments. Total transmission is affected by reflection from both sides (about 4% on each side; Fresnel equation). A 2% increase in transmittance of one-side coated glass roughly halves the one-side reflection. **f**, Contact angles for glass and the nanocoatings obtained by protocols 3 and 4. Individual data are shown as grey circles; the bar and error bars indicate mean ± s.d.; *n* = 6 independent experiments. The statistical significance of differences from glass was assessed using a two-tailed *t*-test. For each surface, a representative photo of a water droplet is shown below the plot.

increased light reflection but decreased adhesion of corneal surfaces (Fig. 2d, e). The degradation of nanostructures caused by overexpression of CG5326 and downregulation of Retinin resulted in loss of anti-reflective and anti-adhesive functions (Fig. 2d, e).

Induced folding of Retinin by waxes

Retinin and wax lipid(s) are the two components that jointly regulate the formation and type of corneal nanostructure in *Drosophila* flies. The interaction seen in genetic manipulations corresponds to what is expected if Retinin is the Turing activator and wax is the inhibitor. The Turing model also predicts physical interaction between the two morphogens. To assess this possibility, we established robust low-cost production of recombinant Retinin from bacteria. Retinin (but not similarly sized, unrelated protein Dhit[31]) efficiently bound a set of commercial waxes in a lipid-strip assay (Fig. 3a) and interacted with lanolin wax in a flotation assay (Fig. 3b).

Because Retinin is poorly characterized, we used circular dichroism to assess the secondary structure of this protein, finding that Retinin consists mainly of random coil structures (Fig. 3c, blue). Addition of lanolin in a concentration-dependent manner changed the circular-dichroism spectrum, indicating substantial formation of ordered α-helixes and/or β-sheets (Fig. 3c). Such changes in the chiral properties of a protein usually indicate folding[32,33]. To confirm this, we used Thioflavin T (ThT) fluorescence. ThT binds to proteins that contain β-sheets of minimally four β-strands, inducing fluorescence enhancement of the fluorophore[34]. As controls, we used the β-sheet-containing protein Dhit[31], and BSA as an α-helix-only protein. In the absence of lipids, Retinin showed a complete lack of ThT binding, just as BSA did and in contrast to Dhit (Fig. 3d). After increasing the concentrations of lanolin in the admixture with Retinin, ThT fluorescence gradually increased, while fluorescence in lanolin admixtures with the two control proteins remained unchanged (Fig. 3d).

Thus, we observe that Retinin interacts directly with waxes, and two independent methods reveal that this interaction changes the Retinin conformation from an unstructured protein to a protein with notable secondary structures (Fig. 3e). For several lipid-binding proteins, interaction with lipids is necessary for proper and complete protein folding[35]. We suspect that the induced folding of Retinin on interaction with corneal waxes, which changes the properties of Retinin, is an important step in the self-assembly of the nanostructures.

Engineering of bioinspired nanocoatings

Having demonstrated the physical interaction, we created nanostructures from recombinant Retinin (Supplementary Fig. 17a) and commercial waxes. Such artificial reconstruction qualitatively confirms the identities of the building blocks and their mode of interaction. Mixtures of the two components efficiently coated glass surfaces with nanostructures, whereas Retinin alone, wax alone or control protein admixtures were inefficient (Fig. 4a, Supplementary Fig. 17b, c). The bio-inspired nanocoatings were stable, withstanding intensive and lengthy washing (Supplementary Fig. 17d).

We created a set of artificial nanocoatings by varying different parameters of the admixture. One parameter was the protein-to-wax ratio (Fig. 4a, Supplementary Fig. 17c). Nanocoatings with small isolated protrusions (similar to the degraded corneal nanocoatings, Fig. 2c) were produced at a Retinin-to-lanolin ratio of 3:17; nanocoatings with higher and more tightly packed protrusions were produced at a ratio of 4:16. Further increases in the Retinin load merged individual protrusions into maze-like structures (5:15 ratio) and into filled fields interspersed with dimple-like depressions (15:5 ratio, Fig. 4a). Similar types of nanocoating occur naturally in different insect species[4] (compare Fig. 4a and Supplementary Fig. 17e; the resemblance also noticeable by Fourier transformation to measure the compaction and ordering of the protrusions[2,15]). We were unable to achieve the degree of order seen in the *Drosophila* corneal nanostructures in our reconstituted nanocoatings, providing a direction for future studies.

Another parameter is the type of wax in the admixture. Waxes have different viscosities (related to their melting points, T_m) and thus different diffusion coefficients—the key characteristic in them acting as a Turing inhibitor. Admixing Retinin with carnauba wax ($T_m \approx 82$ °C), beeswax ($T_m \approx 62$ °C) or lanolin ($T_m \approx 38$ °C) produced nanocoatings with different patterns and progressively broader unit size (Supplementary Fig. 17f, g), in remarkable agreement with Turing modelling (Supplementary Fig. 8).

We generated different protocols to diversify our bio-inspired nanocoatings, producing a large variety of coatings with different dimensions and patterns (Fig. 4d). Further means for diversification may include varying the surface material or pH of the admixture, using Retinin-like proteins from other insects, and using Retinin modifications that permit post-assembly metal capturing, antibody or enzyme binding, enzymatic modifications, and so on.

Diverse nanocoatings result in diverse functions. When studying nanostructures of insect corneas, we looked at anti-reflectance (which increases light transmittance) and anti-adhesion (which is related to hydrophobicity[36]). By studying these functions in our bio-inspired artificial nanocoatings, we find that some of the nanocoatings halve light reflectance from glass across the visible spectrum (Fig. 4b, e). We also find that others have anti-adhesive properties, depending on their type, composition and the preparation mode (Fig. 4c). Regarding the liquid-handling properties, nanocoatings ranging from hydrophobic to hydrophilic could be produced (Fig. 4f); both extremes could have useful applications. The nanocoatings with small isolated protrusions (similar to the degraded corneal nanocoatings, Fig. 2c) produced at a Retinin-to-lanolin ratio of 3:17 (Fig. 4a) show large deviations in the individual adhesion-force measurements (Fig. 4c), probably because of their topography. The large spaces between the isolated nanoprotrusions are strongly adhesive; by contrast, the adhesion force drops when the force-measuring cantilever touches the protrusions. In this regard, the isolated nipple-type structures can be viewed as transitory between the highly adhesive nanocoating-lacking surface and the fully functional nanocoatings of the maze-type pattern.

There are numerous technological applications of artificial nanocoatings, including in the energy, electronics, automotive, marine, aerospace and medical-devices industries[2]. According to rough estimates, the profit in the global market of materials with nanostructured surfaces will reach US$14.2 billion by 2027[37]. The use of natural biodegradable materials provides an eco-friendly alternative to current methods for industrial micro- and nanostructuring. Nanostructures are already produced in nature in an economical manner, exemplified by the moth-eye nanostructures. Our work has identified how this occurs and translates this knowledge into eco-friendly technological applications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2707-9.

1. Nalwa, H. S. *Handbook of Nanostructured Biomaterials and their Applications in Nanobiotechnology* (American Scientific, 2005).
2. Kryuchkov, M., Blagodatski, A., Cherepanov, V. & Katanaev, V. L. in *Functional Surfaces in Biology III: Diversity of the Physical Phenomena* (eds Gorb, S. N. & Gorb, E. V.) 29–52 (Springer, 2017).
3. Peisker, H. & Gorb, S. N. Always on the bright side of life: anti-adhesive properties of insect ommatidia grating. *J. Exp. Biol.* **213**, 3457–3462 (2010).
4. Blagodatski, A., Sergeev, A., Kryuchkov, M., Lopatina, Y. & Katanaev, V. L. Diverse set of Turing nanopatterns coat corneae across insect lineages. *Proc. Natl Acad. Sci. USA* **112**, 10750–10755 (2015).
5. Turing, A. M. The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. B* **237**, 37–72 (1952).
6. Kondo, S. & Miura, T. Reaction-diffusion model as a framework for understanding biological pattern formation. *Science* **329**, 1616–1620 (2010).
7. Bhushan, B. *Springer Handbook of Nanotechnology* 4th edn (Springer, 2017).
8. Adams, M. D. et al. The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
9. Drosophila 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
10. Büscher, T. H., Kryuchkov, M., Katanaev, V. L. & Gorb, S. N. Versatility of Turing patterns potentiates rapid evolution in tarsal attachment microstructures of stick and leaf insects (Phasmatodea). *J. R. Soc. Interface* **15**, 20180281 (2018).

11. Gemne, G. ontogenesis of corneal surface ultrastructure in nocturnal Lepidoptera. *Philos. Trans. R. Soc. Lond. B* **262**, 343–363 (1971).
12. Murray, J. D. *Mathematical Biology II: Spatial Models and Biomedical Applications* (Springer, 2001).
13. Markow, T. A. & O'Grady, P. M. *Drosophila* biology in the genomic age. *Genetics* **177**, 1269–1276 (2007).
14. Bernhard, C. G. & Miller, W. H. A corneal nipple pattern in insect compound eyes. *Acta Physiol. Scand*. **56**, 385–386 (1962).
15. Kryuchkov, M. et al. analysis of micro- and nano-structures of the corneal surface of *Drosophila* and its mutants by atomic force microscopy and optical diffraction. *PLoS One* **6**, e22237 (2011).
16. Kryuchkov, M., Lehmann, J., Schaab, J., Fiebig, M. & Katanaev, V. L. Antireflective nanocoatings for UV-sensation: the case of predatory owlfly insects. *J. Nanobiotechnology* **15**, 52 (2017).
17. Stark, W. S. & Wasserman, G. S. Transient and receptor potentials in the electroretinogram of *Drosophila*. *Vision Res*. **12**, 1771–1775 (1972).
18. Anderson, M. S. & Gaimari, S. D. Raman-atomic force microscopy of the ommatidial surfaces of Dipteran compound eyes. *J. Struct. Biol*. **142**, 364–368 (2003).
19. Chandran, R., Williams, L., Hung, A., Nowlin, K. & LaJeunesse, D. SEM characterization of anatomical variation in chitin organization in insect and arthropod cuticles. *Micron* **82**, 74–85 (2016).
20. Kaya, M., Sargin, I., Al-Jaf, I., Erdogan, S. & Arslan, G. Characteristics of corneal lens chitin in dragonfly compound eyes. *Int. J. Biol. Macromol*. **89**, 54–61 (2016).
21. Locke, M. The Wigglesworth lecture: insects for studying fundamental problems in biology. *J. Insect Physiol*. **47**, 495–507 (2001).
22. Nickerl, J., Tsurkan, M., Hensel, R., Neinhuis, C. & Werner, C. The multi-layered protective cuticle of Collembola: a chemical analysis. *J. R. Soc. Interface* **11**, 20140619 (2014).
23. Kryuchkov, M. et al. Alternative moth-eye nanostructures: antireflective properties and composition of dimpled corneal nanocoatings in silk-moth ancestors. *J. Nanobiotechnology* **15**, 61 (2017).
24. Kim, E. et al. Characterization of the *Drosophila melanogaster* retinin gene encoding a cornea-specific protein. *Insect Mol. Biol*. **17**, 537–543 (2008).
25. Komori, N., Usukura, J. & Matsumoto, H. Drosocrystallin, a major 52 kDa glycoprotein of the Drosophila melanogaster corneal lens. Purification, biochemical characterization, and subcellular localization. *J. Cell Sci*. **102**, 191–201 (1992).
26. Karouzou, M. V. et al. *Drosophila* cuticular proteins with the R&R Consensus: annotation and classification with a new tool for discriminating RR-1 and RR-2 sequences. *Insect Biochem. Mol. Biol*. **37**, 754–760 (2007).
27. Stahl, A. L., Charlton-Perkins, M., Buschbeck, E. K. & Cook, T. A. The cuticular nature of corneal lenses in *Drosophila melanogaster*. *Dev. Genes Evol*. **227**, 271–278 (2017).
28. Cheng, J. B. & Russell, D. W. Mammalian wax biosynthesis. I. Identification of two fatty acyl-coenzyme A reductases with different substrate specificities and tissue distributions. *J. Biol. Chem*. **279**, 37789–37797 (2004).
29. Cheng, J. B. & Russell, D. W. Mammalian wax biosynthesis. II. Expression cloning of wax synthase cDNAs encoding a member of the acyltransferase enzyme family. *J. Biol. Chem*. **279**, 37798–37807 (2004).
30. Kunst, L. & Samuels, A. L. Biosynthesis and secretion of plant cuticular wax. *Prog. Lipid Res*. **42**, 51–80 (2003).
31. Lin, C. et al. Double suppression of the Gα protein activity by RGS proteins. *Mol. Cell* **53**, 663–671 (2014).
32. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta Proteins Proteom*. **1751**, 119–139 (2005).
33. Clarke, D. T. in *Protein Folding, Misfolding, and Disease: Methods and Protocols* (eds Hill, A. F. et al.) 59–72 (Humana, 2011).
34. Biancalana, M. & Koide, S. Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta Proteins Proteom*. **1804**, 1405–1412 (2010).
35. Chandra, S., Chen, X., Rizo, J., Jahn, R. & Sudhof, T. C. A broken α-helix in folded α-synuclein. *J. Biol. Chem*. **278**, 15313–15318 (2003).
36. van der Werf, K. O., Putman, C. A. J., Degrooth, B. G. & Greve, J. Adhesion force imaging in air and liquid by adhesion mode atomic-force microscopy. *Appl. Phys. Lett*. **65**, 1195–1197 (1994).
37. Global Industry Analysts *Nanocoatings — Global Market Trajectory and Analysis* https://researchandmarkets.com/reports/4721438/nanocoatings-global-market-trajectory-and (2020).

Methods

Drosophila cultivation and genetic manipulations

The following sources of different *D. melanogaster* lines were used in this study: the Bloomington *Drosophila* stock centre, Oregon-R-C (as a wild-type control), *GMR-Gal4* (driving the expression in all post-mitotic eye cells[38]), *spa-Gal4* (expressed in the lens-secreting cone cells of the ommatidia[38]), φX-22A (for germ-line transformation[39]) and *CG1942*^{f03896} (stock #18707, acyltransferase mutant); the Vienna *Drosophila* Resource Center, *UAS-RNAi-Crys* (line #37736 GD), *UAS-RNAi-retinin* (102711 KK), *UAS-RNAi-Cpr72Ec* (29452 GD), *UAS-RNAi-CG2781* (48139 GD), *UAS-RNAi-CG31523* (45226 GD), *UAS-RNAi-Baldspot* (47521 GD), *UAS-RNAi-CG33110* (29689 GD), *UAS-RNAi-CG5326* (47681 GD) and *UAS-RNAi-CG31522* (37329 GD); and the FlyORF (Zurich ORFeome Project), *UAS-CG5326* (stock #1245). *D. melanogaster* fruit flies were raised at 25 °C in pursuance of the conventional fly husbandry guidelines[40].

cDNA of *Retinin* (DGRC RH08687) was digested by EagI and Bsp120I sites and subcloned into the NotI site of the pUASTattB plasmid, after which the constructed transgenes were sequenced using vector-specific primers: gtaaccagcaaccaagta (forward) and gtccaattatgtcacacc (reverse). The constructs were used for generation of the transgenic *UAS-Retinin* line through site-specific germ-line transformation of φX-22A line with attP-landing site on the chromosome arm 2L[39].

The following non-*melanogaster Drosophila* stocks were obtained from the University of California, San Diego, *Drosophila* Stock Center: *D. pseudoobscura* (stock #14011-0121.00), *D. simulans* (14021-0251.001), *D. virilis* (15010-1051.00), *D. erecta* (14021-0224.00), *D. ananassae* (14024-0371.00), *D. mojavensis* (15081-1352.01), *D.yakuba* (14021-0261.00), *D. persimilis* (14011-0111.01), *D. sechellia* (14021-0248.03), *D. suzukii* (14023-0311.00), *D. willistoni* (14030-0811.00), *D. busckii* (13000-0081.00) and *D. grimshawi* (15287-2541.00).

Preparation and analysis of corneal and retinal samples

Corneal and retinal samples were prepared by cutting off, with a scalpel, the eyes from the heads of mature-adult guillotined *Drosophila*. The retinal material was removed from immobilized samples into a drop of water by washing and very gentle and scrupulous scratching. After separation, the corneal material was further washed three times in water. The same material was analysed by AFM. Corneal and retinal samples were collected from material of 20 eyes. The samples were boiled for 15 min in the sample buffer (62.5 mM Tris-HCl, pH 6.8; 10% glycerol; 2% SDS; 1% β-mercaptoethanol; trace of bromophenol blue) before separation by 15% SDS–PAGE; the remaining hard corneal material was assessed for Supplementary Fig. 10b by AFM.

Mass spectrometry

Following the SDS–PAGE, bands corresponding to the major and minor corneal proteins were excised, and in-gel trypsin digestion and mass-spectrometry was performed by the Proteomics Facility of the University of Konstanz. The identification of Retinin in non-*melanogaster* species was performed by the Protein Analysis Facility of the University of Lausanne.

Quantification of Retinin levels

Levels of Retinin in Coomassie-stained SDS–PAGE or western blots were quantified by using ImageJ, a free, Java-based image-processing package (https://rsb.info.nih.gov/ij/). For the Coomassie-stained SDS–PAGE, the percentage of the area under the peak, corresponding to Retinin, compared to the area under whole graph was measured. For western blots, levels of Retinin protein was normalized by Tubulin levels, and divided by the normalized result for control genotype (Oregon R-C).

AFM

For AFM, corneal samples prepared as described above were attached to a coverslip by double-sided bonding tape. Microscopy was performed by the NTegra-Prima microscope (NT-MDT) and NSG 11 long (NT-MDT) as a cantilever in contact mode. Artificial nanocoatings were measured using the semi-contact procedure.

Adhesion-force measurement

The same samples that were investigated by AFM were subjected to adhesion-force measurements, performed with the NTegra-Prima microscope (NT-MDT) using the NSG 11 long (NT-MDT) as a cantilever. Measurements were performed on top of ommatidia. Each measurement averages over 300 individual data points across the corneal surface. To exclude incorrectly measured data points, a cut-off was applied for the data with values less than 1 nA. The same protocol was used for artificial nanocoatings.

Reflectance and transmittance measurements

Owing to the impossibility of direct measurement of light transmittance through the cornea of different *Drosophila* species' and mutants' eyes, reflectance was measured, using the JASCO MSV-370 microspectrophotometer. To avoid chromatic aberrations, a non-dispersive Schwarzschild objective was used.

The measurement position on the sample was defined with an aperture (300 µm × 300 µm). The spectral region from ultraviolet (250 nm) to near-infrared (750 nm) was measured. The average spectrum of three repetitions for *D. virilis* and *D. busckii* was used as the baseline for all graphs. For the artificial nanocoatings on glass surface, direct measurements of transmittance were performed with the same settings as for the reflection measurements. The transmittance through air was used as baseline.

Wettability test

In the case of artificial nanocoatings, we can identify the hydrophobicity of the sample directly, by measuring the contact angle between the sample and the surfaces of water drops. A 3-µl water droplet was carefully placed on top of the sample surface. The images were captured with a digital camera and analysed with the Gwyddion software[41], measuring the contact angles from both sides of the droplet.

Nipple parameter quantification and Fourier analysis

The Gwyddion software[41] was used for visualization, cross-section and Fourier analysis. Average height was calculated using an in-house MATLAB script after measuring three vertical cross-sections for each protrusion. Secant planes were defined by pairs of vertices of neighbouring protrusions. The script is provided as Supplementary Methods in Supplementary Information.

Retinin purification and generation of polyclonal rabbit anti-Retinin antibodies

The *Retinin* cDNA without the sequence corresponding to the signal peptide (amino acids 1–21) was amplified from the pUASTattB-retinin plasmid with following primers: forward, ctgtatacatatgagaggatctcaccatcaccatcaccatgccagcttggagtggccctc; reverse, ctgttgactcgagccttagttgcggatgagaaccactcg. The forward primer encompasses the RGSHis-tag coding sequence, adding the tag to the N terminus of Retinin. The PCR product was subcloned into the NdeI and XhoI sites of pET23b and the resulting plasmid was transformed into the Rosetta-gami (Novagen) *Escherichia coli* strain for recombinant expression on induction by IPTG. The bacterial mass was lysed by French press (Constant Systems). RGSHis-Retinin was purified using the HisPur Ni-NTA resin (ThermoFisher Scientific) following the manufacturer's recommendations and used to prepare polyclonal rabbit anti-Retinin antiserum by Eurogentec. RGSHis-Dhit[31] was purified in parallel for control experiments. In western blots, the anti-Retinin antiserum was used at 1:200 dilution. Antibodies to α-tubulin (GTX102079, Lucerna-Chem) were used to probe for the loading control. To identify RGSHis-Retinin in control experiments, the antibodies to the RGSHis-tag (QIAexpress, QIAGEN) were used.

Preparation of wax emulsions

4 g of wax (paraffin, beeswax, carnauba wax #1 (Aldrich Chemistry) or lanolin (Sigma)) was added to tubes with 40 ml 10% SDS solutions in water, and sonicated in a water bath (AL 04-04, Advantage-Lab) for 2 h at a temperature of 80 °C. After subsequent 24-h incubation at room temperature, the upper part enriched in wax nanodrops was diluted in 1×PBS (in water for zeta potential measurement) tenfold and further incubated for 48 h at room temperature. The upper part of the resulting mixture, enriched in wax drops bigger than 500 nm, was discarded. The lower part was dissolved in 1×PBS (in water for zeta potential measurement) to $OD_{600} = 0.5$ (roughly tenfold dilution). These emulsions are stable at room temperature for one year.

Lipid-strip assay to monitor specific interaction of Retinin with waxes

Drops with 3 µl of a wax emulsion were dried on nitrocellulose at 50 °C, and blocked with 3.5% fat-free milk-powder solution in 1×PBS overnight at 4 °C. Next, the nitrocellulose was cut into two parts, the first part to be incubated with 3.5% fat-free milk-powder solution in 1×PBS with recombinant Retinin (1.2 µg ml^{-1}) and the second with 3.5% fat-free milk-powder solution in PBS with the Dhit protein (1.2 µg ml^{-1}) for 1 h at room temperature. Both parts were washed three times by 1×PBS and blocked with 3.5% fat-free milk powder in PBS for 1 h. Antibodies to RGSHis-tag (QIAexpress, QIAGEN) were used following the usual western blotting conditions.

Flotation assay

100 µl of RGSHis-Retinin (0.2 µg µl^{-1}) was incubated with 15 µl of 2.5× lanoline emulsion (fourfold dilution on the last step of emulsion preparation) for 30 min at room temperature. Another 100 µl was incubated for 30 min at room temperature with 15 µl of 0.25% SDS in PBS as a control. In these solutions, 100 µl of 60% sucrose in PBS was added to yield a 30% final sucrose concentration. They were then mixed gently to give a homogenous solution but without disrupting the interaction. These solutions were transferred into Mickrofuge tubes (P10430MPI, Beckman), overlayed by 250 µl of 25% sucrose and 50 µl of PBS buffer on top without disturbing the layers. They were then centrifuged at 100,000*g* for 3 h at 4 °C (corresponding to 40,000 rpm on a Sorvall S45A rotor). The fractions were collected, starting from the top of the tube (100 µl top fraction, 200 µl middle fraction, 200 µl bottom fraction). Antibodies to RGSHis-tag (QIAexpress, QIAGEN) were used following the usual western blotting conditions.

Circular dichroism

Measurements were performed in the Department of Organic Chemistry of the University of Geneva, using the Jasco J-815 circular-dichroism spectrometer with strain-free QS quartz 1-mm path-length cuvettes. Retinin was used at 0.07 mg ml^{-1} in 50 mM potassium phosphate buffer, pH 7.5.

ThT fluorescence measurements

ThT (Sigma, catalogue #T3516) was dissolved in 1×PBS and filtered through a 0.2-µm syringe filter. Final concentrations were 30 µM ThT, 0.04 mg ml^{-1} proteins and 0.004% SDS. Before measurement, ready solutions were incubated for 10 min at room temperature. The ThT fluorescence levels were measured at room temperature by the Infinite M Plex plate reader (Tecan) from the plate (microplate 96 well, greiner bio-one) bottom with excitation of 450 nm and emission of 490 nm in three repetitions. For the ThT fluorescence kinetics experiments, a mixture of ThT and lanolin was added to the protein solutions (Retinin or Dhit), with the following resulting concentrations: 30 µM ThT, about 0.3 mM (0.2 mg ml^{-1}) lanolin and 1 µM protein.

In vitro coating

20 µl of a mixture of Retinin (0.6 mg ml^{-1} in PBS) and the lanolin wax emulsion at different proportions (Fig. 4a) was distributed evenly on a 1-cm^2 area of a glass cover slip and permitted to dry out gradually at room temperature for 20 min at a humidity of 50%–60%, rinsed in water, and re-dried. This process was repeated twice. For Supplementary Fig. 17f, the same protocol was used with different wax emulsions. For Supplementary Fig. 17c, the single-coating protocol, otherwise identical to that for Fig. 4a, was used.

For protocols 1 and 2 (Fig. 4d), only wax emulsions (lanolin for protocol 1 and carnauba wax emulsion with pH = 9 for protocol 2) were used for the first coating step, as above. In the second coating step, 20 µl of Retinin alone (0.6 mg ml^{-1}) was used. For protocol 3, 5 µl of Retinin (0.6 mg ml^{-1} in PBS) and 15 µl of the carnauba wax emulsion were used for the first and second steps. For protocol 4, 5 µl of Retinin (0.6 mg ml^{-1} in PBS) and 15 µl of the lanolin wax emulsion were used for first step, followed by application of 20 µl of Retinin alone (0.6 mg ml^{-1}) as the second step.

AFM, reflectance and wetting measurements were performed on the resultant surfaces as described above for the corneal surfaces.

Zeta potential measurement

Measurement was performed on ZetaSizer Nanoseries (Malvern) analyser, using DTS 1070 cells (Malvern) according to the manufacturer's protocol. Lanolin in water emulsion was used at 0.3 mg ml^{-1} and Retinin at 0.006 mg ml^{-1}, with 0.01× final concentration of PBS.

Block-copolymer modelling

Simulations were made by using an object-oriented framework (PSim, Tech-X) that simulates phase morphologies of dense block copolymers melt systems (https://www.txcorp.com/psim), in bulk linear diblock and diblock + homopolymer mixture simulation models, with the X, Y and Z dimensions in pixels, XN = 128, YN = 128, ZN = 1, the Flory–Huggins interaction parameter, $\chi N = 18$ and the ratio of the length of block A to the length of the whole copolymer length, fA from 0.3 to 0.5 (0.5 for Supplementary Fig. 7a, c; 0.3 for Supplementary Fig. 7d).

3D Turing modelling

For this simulation, the software Ready—a cross-platform implementation of various reaction–diffusion systems (https://github.com/GollyGang/ready)—was used, with two different scripts for in vivo (Supplementary Figs. 2, 3, 4a, 5) and in vitro (Supplementary Fig. 8) coating simulations. The following parameters were used: the activators' degradation, da = 0.03, the activators' diffusion, Da = 0.02, the activators' secretion, sa, from 0.05 to 6.4 (1 for Supplementary Figs. 4a, 5b, 8; 1.2 for Supplementary Figs. 2, 3), the activators' autoactivation, aa = 0.067, the activators' inhibition, ia = −0.08, the inhibitors' degradation, db = 0.08, the inhibitors' diffusion, Db = 0.45 (0.35–1.00 for Supplementary Fig. 7), the inhibitors' secretion, sb, from 0.05 to 12.8 (1 for Supplementary Figs. 2, 3, 4a, 5a, 8), the inhibitors' activation, ab = 0.09 and the inhibitors' autoinhibition, ib = −0.07. The scripts for in vitro and in vivo nanocoating simulations are available at https://github.com/GollyGang/ready/blob/gh-pages/Patterns/Kryuchkov2020/*Drosophila*_corneal_nanocoatings.vti.

Collection of insects

The samples of *Dictyoptera aurora* and *Chalcophora mariana* were provided by Vladimir Savitsky from the Faculty of Biology, Lomonosov Moscow State University. *Anax imperator* and *Liposcelis* sp. were collected in Vaud canton, Switzerland.

Data availability

The data that support the findings of this study are available within the paper and its Supplementary Information. Source data are provided with this paper.

Code availability

The Supplementary Information contains the Matlab script used.

38. Katanaev, V. L. & Kryuchkov, M. V. The eye of Drosophila as a model system for studying intracellular signaling in ontogenesis and pathogenesis. *Biochemistry (Mosc.)* **76**, 1556–1581 (2011).
39. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific ϕC31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
40. Roberts, D. B. *Drosophila: A Practical Approach* (Oxford Univ. Press, 1998).
41. Nečas, D. & Klapetek, P. Gwyddion: an open-source software for SPM data analysis. *Cent. Eur. J. Phys.* **10**, 181–188 (2012).

Acknowledgements We thank A. Koval for MATLAB programming and members of the Katanaev lab for reading the manuscript.

Author contributions M.K. performed most experiments; O.B. performed a set of genetics experiments; J.L. participated in, and M.F. supervised, the physical measurements (AFM, reflectance, and so on); V.L.K. designed the work and guided the experiments. All authors participated in writing the paper.

Competing interests M.K. and V.L.K. are inventors (University of Lausanne) on a patent application (EP18175103.3) for artificial insect-like nanocoatings. Other authors do not have any competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2707-9.

Correspondence and requests for materials should be addressed to V.L.K.

Peer review information *Nature* thanks Shigeru Kondo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Population genomics of the Viking world

https://doi.org/10.1038/s41586-020-2688-8

Received: 12 July 2019

Accepted: 21 May 2020

Ashot Margaryan[1,2,3,71], Daniel J. Lawson[4,5,71], Martin Sikora[1,71], Fernando Racimo[1,71],
Simon Rasmussen[6], Ida Moltke[7], Lara M. Cassidy[8], Emil Jørsboe[7,9], Andrés Ingason[1,10,11],
Mikkel W. Pedersen[1], Thorfinn Korneliussen[1,12], Helene Wilhelmson[13,14], Magdalena M. Buś[15],
Peter de Barros Damgaard[1], Rui Martiniano[16], Gabriel Renaud[1,17], Claude Bhérer[18],
J. Víctor Moreno-Mayar[1,19], Anna K. Fotakis[3], Marie Allen[15], Raili Allmäe[20], Martyna Molak[21],
Enrico Cappellini[3], Gabriele Scorrano[3], Hugh McColl[1], Alexandra Buzhilova[22], Allison Fox[23],
Anders Albrechtsen[7], Berit Schütz[24], Birgitte Skar[25], Caroline Arcini[26], Ceri Falys[27],
Charlotte Hedenstierna Jonson[28], Dariusz Błaszczyk[29], Denis Pezhemsky[22],
Gordon Turner-Walker[30], Hildur Gestsdóttir[31], Inge Lundstrøm[3], Ingrid Gustin[13],
Ingrid Mainland[32], Inna Potekhina[33], Italo M. Muntoni[34], Jade Cheng[1], Jesper Stenderup[1],
Jilong Ma[1], Julie Gibson[32], Jüri Peets[20], Jörgen Gustafsson[35], Katrine H. Iversen[6,17],
Linzi Simpson[36], Lisa Strand[25], Louise Loe[37], Maeve Sikora[38], Marek Florek[39], Maria Vretemark[40],
Mark Redknap[41], Monika Bajka[42], Tamara Pushkina[43], Morten Søvsø[44], Natalia Grigoreva[45],
Tom Christensen[46], Ole Kastholm[47], Otto Uldum[48], Pasquale Favia[49], Per Holck[50], Sabine Sten[51],
Símun V. Arge[52], Sturla Ellingvåg[1], Vayacheslav Moiseyev[53], Wiesław Bogdanowicz[21],
Yvonne Magnusson[54], Ludovic Orlando[55], Peter Pentz[46], Mads Dengsø Jessen[46],
Anne Pedersen[46], Mark Collard[56], Daniel G. Bradley[8], Marie Louise Jørkov[57], Jette Arneborg[46,58],
Niels Lynnerup[57], Neil Price[28], M. Thomas P. Gilbert[3,59], Morten E. Allentoft[1,60], Jan Bill[61],
Søren M. Sindbæk[62], Lotte Hedeager[63], Kristian Kristiansen[64], Rasmus Nielsen[1,65,66] ✉,
Thomas Werge[1,10,11,67] ✉ & Eske Willerslev[1,68,69,70] ✉

The maritime expansion of Scandinavian populations during the Viking Age (about AD 750–1050) was a far-flung transformation in world history[1,2]. Here we sequenced the genomes of 442 humans from archaeological sites across Europe and Greenland (to a median depth of about 1×) to understand the global influence of this expansion. We find the Viking period involved gene flow into Scandinavia from the south and east. We observe genetic structure within Scandinavia, with diversity hotspots in the south and restricted gene flow within Scandinavia. We find evidence for a major influx of Danish ancestry into England; a Swedish influx into the Baltic; and Norwegian influx into Ireland, Iceland and Greenland. Additionally, we see substantial ancestry from elsewhere in Europe entering Scandinavia during the Viking Age. Our ancient DNA analysis also revealed that a Viking expedition included close family members. By comparing with modern populations, we find that pigmentation-associated loci have undergone strong population differentiation during the past millennium, and trace positively selected loci—including the lactase-persistence allele of *LCT* and alleles of *ANKA* that are associated with the immune response—in detail. We conclude that the Viking diaspora was characterized by substantial transregional engagement: distinct populations influenced the genomic makeup of different regions of Europe, and Scandinavia experienced increased contact with the rest of the continent.

The events of the Viking Age altered the political, cultural and demographic map of Europe in ways that are evident to this day. Scandinavian diasporas established trade and settlements that stretched from the American continent to the Asian steppe[1]. They exported ideas, technologies, language, beliefs and practices to these lands, developed new socio-political structures and assimilated cultural influences[2].

To explore the genomic history of the Viking Age, we shotgun-sequenced DNA extracted from 442 human remains from archaeological sites dating from the Bronze Age (about 2400 BC) to the Early Modern period (about AD 1600) (Fig. 1, Extended Data Fig. 1). The data from these ancient individuals were analysed together with published data from 3,855 present-day individuals across two reference panels (Supplementary Note 6), and data from 1,118 ancient individuals (Supplementary Table 3).

Scandinavian ancestry and Viking Age origins

Although Viking Age Scandinavian populations shared a common cultural background, there was no common word for Scandinavian identity at this time[1]. Rather than there being a single 'Viking world', a series of interlinked Viking worlds emerged from rapidly growing maritime exploration, trade, war and settlement, following the adoption

---

A list of affiliations appears at the end of the paper.

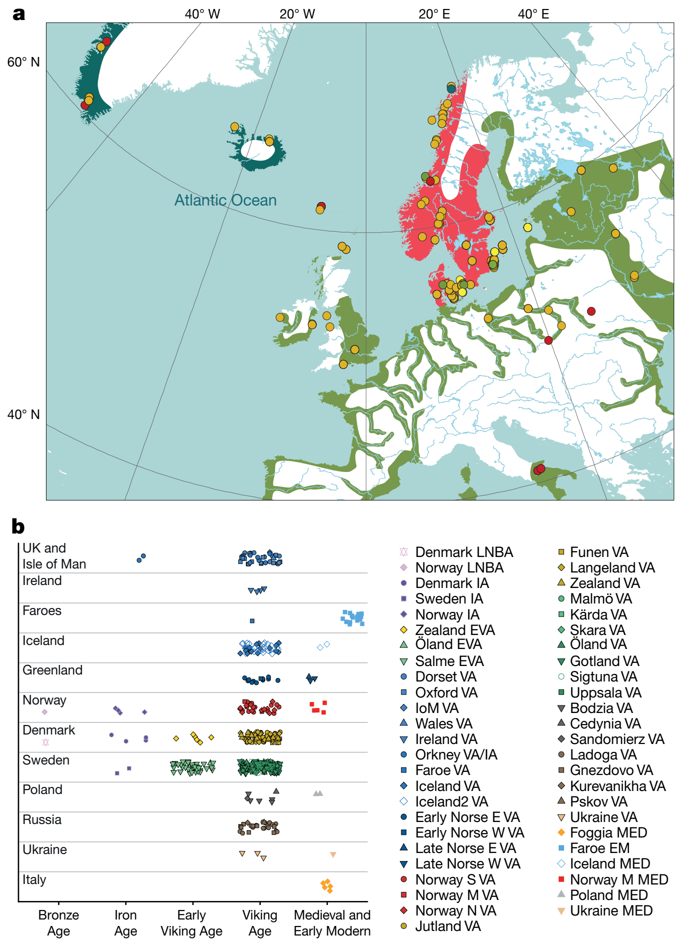


Fig. 1 | Overview of the Viking Age genomic dataset. a, Map of the Viking World from eighth to eleventh centuries AD, showing geographical location and broad age category of sites with ancient samples newly reported in this study. Age categories of sites (circles) are coloured-coded as: dark green, LNBA (2400–500 BC); light green, Iron Age (500 BC–AD 700); yellow, Early Viking Age (AD 700–800); Viking Age (AD 800–1100); Medieval and Early Modern (AD 1100–1600). Red region, area of Viking origins; green region, area of Viking raids, settlement and trade; dark blue region, area of pioneer Viking colonization. **b**, All of the ancient individuals from this study ($n = 442$), and previously published Viking Age samples from Sigtuna[10] and Iceland[18], categorized on the basis of their spatiotemporal origin. The ancient samples are divided into the following five broad categories: Bronze Age (BA), Iron Age (IA), Early Viking Age (EVA), Viking Age (VA), Medieval (MED) and Early Modern (EM). Random jitter has been added along the x axis in each category to aid visualization. LNBA, Late Neolithic and Bronze Age; Norse W, Norse western settlement; Norse E, Norse eastern settlement; Norway S, southern Norway; Norway N, northern Norway; Norway M, middle Norway.

of deep-sea navigation among coastal populations of Scandinavia and the area around the Baltic Sea[3,4]. Thus, it is unclear to what extent the Viking phenomenon refers to people with a recently shared genetic background or how far population changes accompanied the transition from the Iron Age (500 BC–AD 700) to the Viking Age in Scandinavia.

The Viking Age Scandinavian individuals of our study fall broadly within the diversity of ancient European individuals from the Bronze Age and later (Fig. 2, Extended Data Figs. 2, 3, Supplementary Note 8), but with subtle differences among the groups that indicate complex fine-scale structure. For example, many Viking Age individuals from the island of Gotland cluster with Bronze Age individuals from the Baltic region, which indicates mobility across the Baltic Sea (Fig. 2, Extended Data Fig. 3). Using f_4-statistics to contrast genetic affinities with steppe pastoralists and Neolithic farmers, we find that Viking Age individuals from Norway are distributed in a manner similar to that of earlier Iron Age individuals, whereas many Viking Age individuals from Sweden and Denmark show a greater affinity to Neolithic farmers from Anatolia (Extended Data Fig. 4a). Using the qpAdm program, we find that the majority of groups can be modelled as three-way mixtures of hunter-gatherer, farmer and steppe-related ancestry. The three-way model was rejected for some groups from Sweden, Norway and the Baltic region, which could be fit using four-way models that additionally included either Caucasus hunter-gatherer or East-Asian-related ancestry (Extended Data Figs. 4b, c)—the latter of which is consistent with previously documented gene flow from Siberia[5–7].

Investigating genetic continuity with Iron Age groups that are temporally more proximate to the Viking Age Scandinavian populations, we find that most Viking Age groups can be fit using a single Iron Age source and broadly fall into two categories: (i) English Iron Age sources (most of the Viking Age individuals from Denmark, as well as populations of the British Isles) and (ii) Scandinavian Iron Age sources (from Norway, Sweden and the Baltic region) (Extended Data Fig. 5a). Notable exceptions are individuals from Kärda in southern Sweden, for whom only early Medieval Longobard individuals from Hungary can be fit as a single source group ($P > 0.01$) (Extended Data Fig. 5a). Groups with poor one-way fits can be modelled by including either additional north-eastern ancestry (for example, Viking Age individuals from Ladoga) or additional southeastern ancestry (for example, Viking Age individuals from Jutland) (Extended Data Fig. 5b). Overall, our analyses suggest that the genetic makeup of Viking Age Scandinavian populations largely derives from ancestry of the preceding Iron Age populations—but these analyses also reveal subtle differences in ancestry and gene flow from both the south and east. These observations are largely consistent with archaeological findings[8,9].

Viking Age genetic structure in Scandinavia

To elucidate the fine-scale population structure of Viking Age Scandinavia, we performed genotype imputation on a subset of 298 individuals with sufficient (>0.5×) coverage (289 from this study, along with 9 previously published individuals[10]) and inferred the genomic segments they shared via identity-by-descent with a reference panel of present-day European individuals ($n = 1{,}464$) (Supplementary Notes 6, 10, 11). Genetic clustering using multidimensional scaling and uniform manifold approximation and projection (UMAP) shows that Viking Age Scandinavian individuals cluster into three groups by geographical origin, with close affinities to their respective present-day counterparts (Fig. 3a, Supplementary Fig. 10.1). Some individuals—particularly those from the island of Gotland in eastern Sweden—have strong affinities with Eastern Europeans; this probably reflects individuals with Baltic ancestry, as clustering with Bronze Age individuals from the Baltic region is evident in the identity-by-state UMAP analysis (Fig. 2b) and through f_4-statistics (Supplementary Fig. 9.1).

We used ChromoPainter[11] and a reference panel enriched with Scandinavian individuals ($n = 1{,}464$) (Supplementary Notes 6, 11) to identify long, shared haplotypes and detect subtle population structure (Supplementary Figs. 11.1–11.10). We find ancestry components in Scandinavia with (inexact and indicative) affinities with present-day populations (Supplementary Fig. 11.11), which we refer to as ‘Danish-like’, ‘Swedish-like’, ‘Norwegian-like’ and ‘North Atlantic-like’ (that is, possible individuals from the British Isles entering Scandinavia). The sampling is heavily structured, so these complex results (Supplementary Fig. 11.12) are visualized over time and space (Fig. 4) using spatial interpolation[12] to account for sampling locations and report significant trends (Supplementary Table 11.2) using linear regression (Supplementary Notes 11, 12).

Norwegian-like and Swedish-like components cluster in Norway and Sweden, respectively, whereas Danish-like and North-Atlantic-like components are widespread (Fig. 4, Supplementary Fig. 11.12,

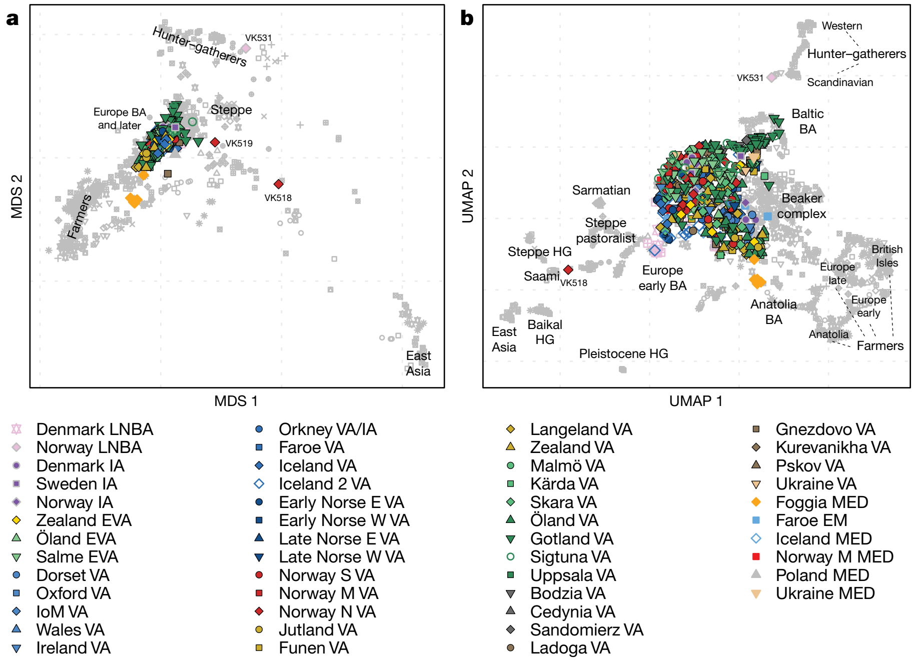


Fig. 2 | Genetic structure of Viking Age samples. a, Multidimensional scaling (MDS) of *n* = 1,305 ancient genomes, on the basis of a pairwise identity-by-state sharing matrix of the Viking Age and other ancient samples (Supplementary Table 3). Outlier individuals with hunter-gatherer (VK531) or Saami-related ancestry (VK518 and VK519) are highlighted. **b**, UMAP analysis of the same dataset as in **a**, with fine-scale ancestry groups highlighted. HG, hunter-gatherer.

Supplementary Table 6). Unexpectedly, Viking Age individuals from Jutland (Denmark) lack Swedish-like and Norwegian-like genetic components (Supplementary Fig. 11.12). We also find that gene flow within Scandinavia was broadly from south to north, dominated by movement from Denmark into Norway and Sweden (Supplementary Table 11.2).

We identified two ancient individuals from northern Norway (designated VK518 and VK519) with affinities to present-day Saami populations in Norway and Sweden. The VK519 individual probably also had Norwegian-like ancestors, which indicates genetic contacts between Saami groups and other Scandinavian populations.

The genetic data are structured by topographical boundaries rather than by the borders of present-day countries. Thus, the southwestern part of Sweden in the Viking Age is genetically more similar to Viking Age populations of Denmark than to those of central mainland Sweden, probably owing to geographical barriers that prevented gene flow.

We quantified genetic diversity using two measures: conditional nucleotide diversity (Supplementary Note 9) and variation in inferred ancestry on the basis of ChromoPainter results (Extended Data Fig. 6, Supplementary Note 11, Supplementary Fig. 11.13). We also visualized this diversity as the spread of individuals on a multidimensional scaling plot based on a pairwise identity-by-state sharing matrix (Fig. 3b).

Diversity varies markedly from the more-homogeneous inland and northern parts of Scandinavia to the diverse Kattegat (eastern Denmark and western Sweden) and Baltic Sea regions, which suggests an important role for these maritime regions in interaction and trade during the Viking Age. On Gotland, there are many more Danish-like and North-Atlantic-like genetic components (as well as an additional 'Finnish-like' ancestry component) than Swedish-like components, which indicates extensive maritime contacts for Gotland during the Viking Age.

Our results for Gotland and Öland agree with archaeological indications that these were important maritime communities from the Roman period (AD 1–400) onwards[13,14]. A similar pattern is observed on the central Danish islands (such as Langeland) but at a lower level. The data indicate that genetic diversity on the islands increased from the early (about eighth century AD) to the late Viking Age (about tenth to eleventh centuries AD), which suggests increasing interregional interaction over time. Evidence for genetic structure within Viking Age Scandinavia[2,4,15–17]—with diversity in cosmopolitan centres such as Skara, and trade-oriented islands such as Gotland—highlight the importance of sea routes during this period.

Viking migrations

Our fine-scale ancestry analyses of genomic data are consistent with patterns documented by historians and archaeologists (Figs. 3, 4, Supplementary Fig. 11.12): eastward movements mainly involved Swedish-like ancestry, whereas individuals with Norwegian-like ancestry travelled to Iceland, Greenland, Ireland and the Isle of Man. The first settlement in Iceland and Greenland also included individuals with North-Atlantic-like ancestry[18,19]. A Danish-like ancestry is seen in present-day England, in accordance with historical records[20], place names[21], surnames[22] and modern genetics[23,24], but Viking Age Danish-like ancestry in the British Isles cannot be distinguished from that of the Angles and Saxons, who migrated in the fifth to sixth centuries AD from Jutland and northern Germany.

Viking Age execution sites in Dorset and Oxford (England) contain North-Atlantic-like ancestry, as well as Danish-like and Norwegian-like ancestries. If these sites represent Viking raiding parties that were defeated and captured[25,26], then these raids were composed of individuals of different origins. This pattern is also suggested by isotopic data from a warrior cemetery in Trelleborg (Denmark)[27]. Similarly, the presence of Danish-like ancestry in an ancient sample from Gnezdovo in present-day Russia indicates that eastern migrations were not entirely composed of Viking individuals from Sweden.

Our results show that 'Viking' identity was not limited to individuals of Scandinavian genetic ancestry. Two individuals from Orkney who were buried in Scandinavian fashion are genetically similar to present-day Irish and Scottish populations, and are probably the first Pictish genomes published (see 'Evidence for Pictish genomes' in Supplementary Note 11, Supplementary Figs. 11.3, 11.12, 11.14, Supplementary Table 6). Two other individuals from Orkney had 50% Scandinavian

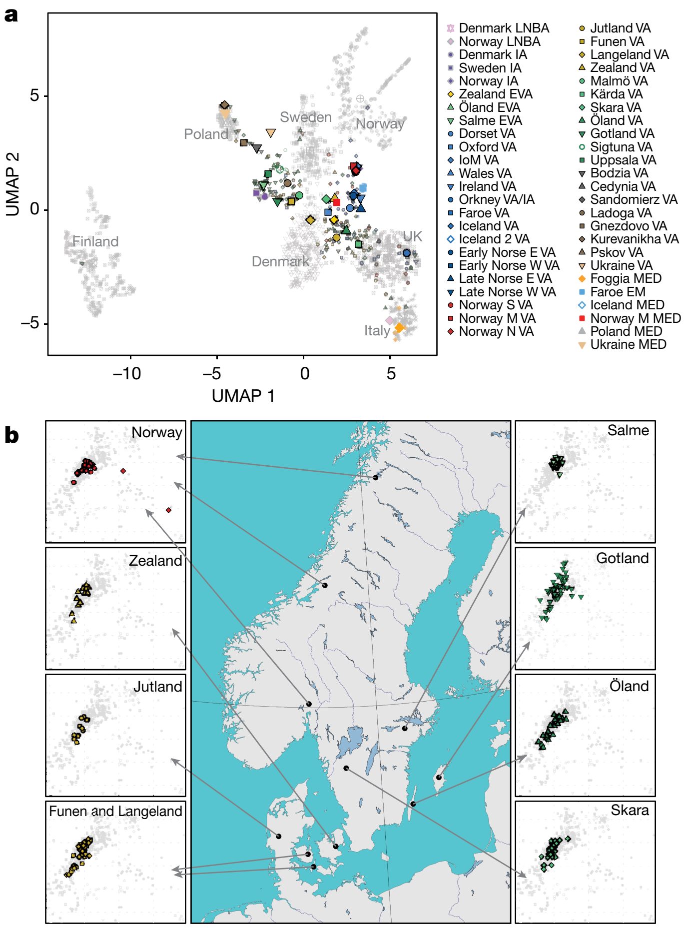


Fig. 3 | Genetic structure and diversity of ancient samples. a, UMAP analysis of $n = 1{,}624$ ancient and modern Scandinavian individuals, on the basis of the first 10 dimensions of MDS using identity-by-descent segments of imputed individuals. Large symbols indicate median coordinates for each group. **b**, Genetic diversity in major populations of the Scandinavian Viking Age. Plots next to the map show MDS analysis on the basis of a pairwise identity-by-state sharing matrix. Norway denotes all the sites from Norway. The scale is identical for all the plots.

ancestry, and five such individuals were found in Scandinavia. This suggests that Pictish populations may have been integrated into Scandinavian culture by the Viking Age.

Viking Age gene flow into Scandinavia

Non-Scandinavian ancestry in samples from Denmark, Norway and Sweden agrees with known trading routes (Supplementary Notes 11, 12): for example, Finnish and Baltic ancestry reached modern Sweden (including Gotland), but is absent in most individuals from Denmark and Norway. By contrast, western regions of Scandinavia received ancestry from the British Isles (Supplementary Notes 11, 12). The first evidence of South European ancestry (>50%) in Scandinavia is during the Viking Age in Denmark (for example, individuals VK365 and VK286 from Bogøvej) and southern Sweden (for example, VK442 and VK350 from Öland, and VK265 from Kärda) (Fig. 4, Supplementary Table 6).

Disappearance from Greenland

From around AD 980 to 1440, southwest Greenland was settled by people of Scandinavian ancestry (probably from Iceland)[28,29]. The fate of these populations in Greenland remains debated, but probable causes of their disappearance are social or economic processes in Europe (for example, political relations within Scandinavia and changed trading systems) and natural processes, including climatic change[29–31].

According to our data, the Greenland Norse populations were an admixture between Scandinavians (mostly from Norway) and individuals from the British Isles, similar to the first settlers of Iceland[18]. We see no evidence of long-term inbreeding in the genomes of Greenlandic Norse individuals, although we have only one high-coverage genome from the later period of occupation of the island (Supplementary Note 10, Supplementary Figs. 10.2, 10.3). This result could favour a relatively brief depopulation scenario, consistent with previous demographic models[32] and archaeological findings. We also find no evidence of ancestry from other populations (Palaeo-Eskimo, Inuit or Native American) in the Greenlandic Norse genomes (Supplementary Fig. 9.4), which accords with the skeletal remains[32]. This suggests that sexual interaction between the Greenland Norse populations and these other groups was absent, or occurred only on a very small scale.

Genetic composition of earliest Viking voyage

Although maritime raiding has been a constant of seafaring cultures for millennia, the Viking Age is partly defined by this activity[33]. However, the exact nature and composition of Viking war parties is unknown[5]. One raiding or diplomatic expedition has left direct archaeological traces: at Salme in Estonia, 41 men from Sweden who died violently were buried in two boats, accompanied by high-status weaponry[34,35]. Importantly, the Salme boat burial predates the first textually documented raid (on Lindisfarne (England) in 793) by nearly half a century.

Kinship analysis of the genomes of 34 individuals from the Salme burial reveals 4 brothers buried side by side and a third-degree relative of 1 of the 4 brothers (Supplementary Note 4). The ancestry profiles of the Salme individuals were similar to one another when compared to the profiles of other burials of the Viking Age (Supplementary Notes 10, 11), which suggests a relatively genetically homogeneous group of people of high status (including close kin).

The five Salme relatives are not the only kin in our dataset; we also identified two pairs of kin in which the related individuals were excavated hundreds of kilometres apart from each other, which markedly illustrates the mobility of individuals during the Viking Age.

Positive selection in northern Europe

We looked for single-nucleotide polymorphisms (SNPs) with allele frequencies that have changed significantly in the last 10,000 years[36,37], beyond what can be explained by temporal changes in ancestry alone (Supplementary Note 14). Extended Data Figure 8a shows the likelihood ratio scores in favour of selection in the entire 10,000-year period (the general scan), the period up to 4,000 years before present (the ancient scan) and the period from 4,000 years before present up to today (the recent scan).

As expected[38,39], the strongest candidates for selection are SNPs near *LCT*, the frequency of which increased after the Bronze Age[40,41]. Our dataset traces the frequency of the lactase-persistence allele (rs4988235) and its evolution since the Bronze Age. Extended Data Figure 8b shows that Viking Age groups had very similar allele frequencies at the *LCT* lactase-persistence SNP to those of present-day northern European populations. Conversely, Bronze Age Scandinavian individuals, as well as individuals from central Europe associated with Corded Ware and Bell Beaker assemblages, have a low frequency of this SNP despite evidence for milk consumption. Our Iron Age samples have intermediate frequencies, which suggests a rise in lactase persistence during this period. The frequency is higher in the Bronze Age of the Baltic Sea region than in Bronze Age Scandinavia, which is consistent with gene flow between the two regions explaining the increasing frequency of lactase persistence in Scandinavia.

Other candidates for selection include previously identified regions, including the one containing the *TLR1*, *TLR6* and *TLR10* genes, the HLA

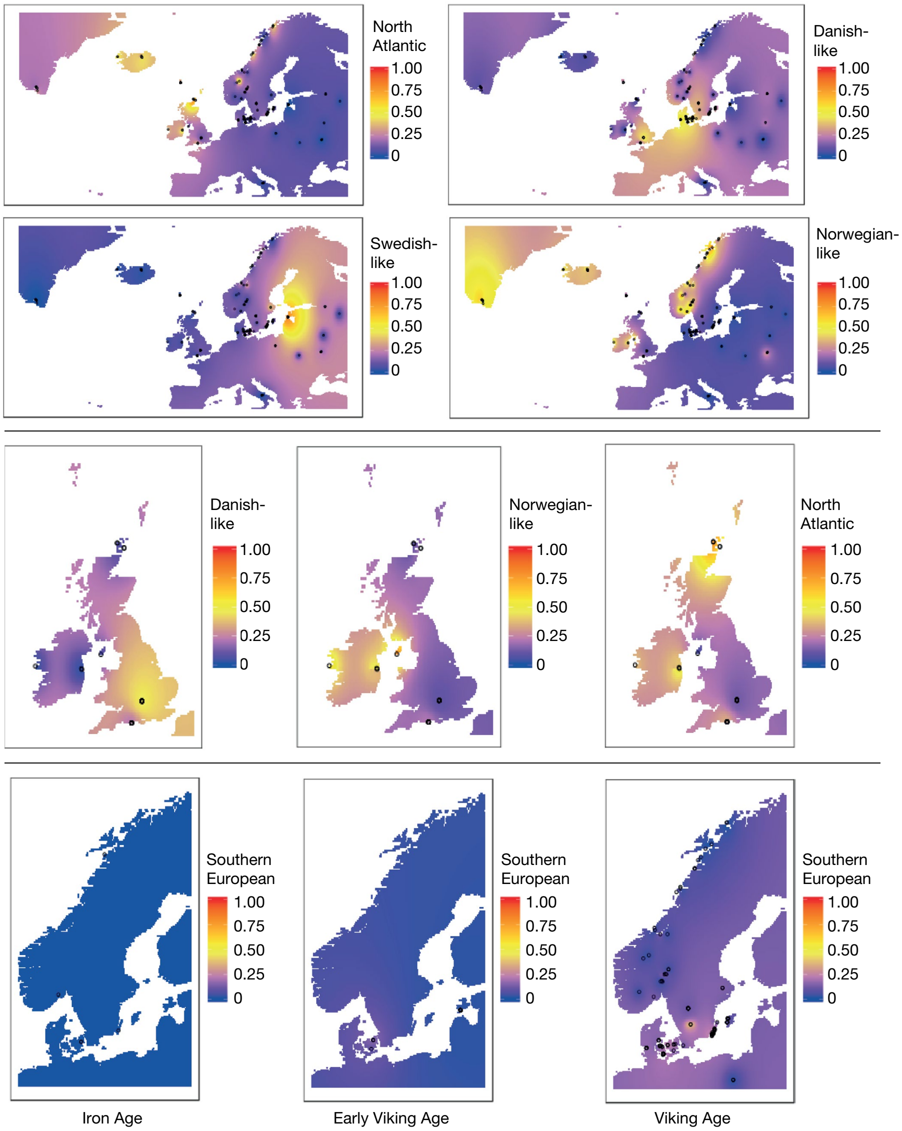


Fig. 4 | Spatiotemporal patterns of Viking and non-Viking ancestry in Europe during the Iron Age, Early Viking Age and Viking Age. We performed inverse distance-weighting interpolation of the ancestry painting proportions of each individual genome on a dense grid of points covering the European continent, to better visualize the distribution of ancestry paintings at different periods (Supplementary Note 12). Top, distinct spheres of influence in the Viking world. Middle, Danish Viking ancestry in southern Britain, Norwegian Viking ancestry in Ireland and Isle of Man and non-Scandinavian ('North Atlantic') ancestry in Orkney, Ireland and southern Britain. Bottom, Late southern European ancestry in southern Scandinavia. The Swedish-like ancestry is the highest in present-day Estonia owing to the ancient samples from the Salme ship burial, which originated from the Mälaren valley of Sweden (according to archaeological sources). $n = 289$ genomes used for interpolation.

region, and the genes *SLC45A2* and *SLC22A4*[41]. We also find additional candidate regions for selection that have associated trajectories that start before the Viking Age, which suggests shared phenotypes between ancient Viking and present-day Scandinavian populations (Supplementary Note 14). These regions include one that overlaps *DCC* and that is implicated in colorectal cancer[42], as well as one that overlaps *AKNA* and is involved in the secondary immune response[43].

Evolution of complex traits in Scandinavia

To search for signals of recent population differentiation at SNP markers associated with complex traits, we compared genotypes of Viking Age individuals with those of a panel of present-day Danish individuals[44]. We obtained summary statistics from 16 well-powered genome-wide association studies through the GWAS ATLAS[45] and tested for a difference in the distribution of polygenic scores between the two groups (Supplementary Note 15). The polygenic scores of Viking Age individuals and present-day Danish individuals differed for three traits: black hair colour ($P = 0.00089$), standing height ($P = 0.019$) and schizophrenia ($P = 0.0096$), although the latter two were not significant after accounting for the number of tests (Extended Data Fig. 7). Currently, we cannot conclude whether the observed differences in allele frequencies are due to selection acting on these alleles between the Viking Age and the present time or to some other factors (such as more ethnic diversity in the Viking Age sample). A binomial test of the number of black hair colour risk alleles at higher frequency in the Viking Age sample and the present-day sample was also significant (65/41; $P = 0.025$), which suggests that the signal is not entirely driven by a few large-effect loci.

Viking genetic legacy in populations today

To test whether present-day Scandinavian populations share increased ancestry with their respective counterparts in the Viking Age, we first computed *D*-statistics of the form *D*(Yoruba (YRI), ancient; present-day population 1, present-day population 2), which measure whether an ancient test individual shares more alleles with present-day population 1 or with present-day population 2. Viking Age individuals shift subtly from Scandinavia towards their present-day counterparts in the distributions of these statistics (Extended Data Fig. 5c, Supplementary Figs. 9.2, 9.3).

We further examined ancient ancestry in present-day populations using fineSTRUCTURE (Supplementary Note 11, Supplementary Fig. 11.14). Within Scandinavia, most present-day populations resemble their Viking Age counterparts. The exception is Swedish-like ancestry, which is present at only 15–30% within Sweden today: one cluster from Sweden is closer to ancient Finnish populations, and a second is more closely related to Danish and Norwegian populations. Danish-like ancestry is now high across the whole region.

Outside of Scandinavia, the genetic legacy of Viking Age populations is consistent—although limited. A small Scandinavian ancestry component is present in Poland (up to 5%). Within the British Isles, it is difficult to assess how much of the Danish-like ancestry is due to pre-existing Anglo-Saxon ancestry, but the Viking Age contribution does not exceed 6% in England (Supplementary Note 11). The genetic effects are stronger in the other direction. Although some North-Atlantic-like individuals in Orkney became culturally Scandinavian, others found themselves in Iceland, Norway and beyond, leaving a genetic legacy that persists today. Present-day Norwegian individuals vary between 12 and 25% in North-Atlantic-like ancestry; this ancestry is more uniformly 10% in Sweden.

Discussion

Our genomic analyses shed light on long-standing questions raised by historical sources and archaeological evidence from the Viking Age. We largely confirm the long-argued movements of Vikings outside Scandinavia: Vikings from present-day Denmark, Norway, and Sweden going to Britain, the islands of the North Atlantic, and sailing east towards the Baltic region and beyond, respectively. However, we also see ancient Swedish-like and Finnish-like ancestry in the westernmost fringes of Europe, and Danish-like ancestry in the east, defying modern historical groupings. It is likely that many such individuals were from communities with mixed ancestries, thrown together by complex trading, raiding and settling networks that crossed cultures and the continent.

During the Viking Age, different parts of Scandinavia were not evenly connected, leading to clear genetic structure in the region. Scandinavia probably comprised a limited number of transport zones and maritime enclaves[46] with active external contacts, and limited external gene flow into the rest of the Scandinavian landmass. Some Viking Age Scandinavian locations are relatively homogeneous—particularly mid-Norway, Jutland and the Atlantic settlements. This contrasts with the strong genetic variation of populous coastal and southern trading communities such as in the islands of Gotland and Öland[47–49]. The high genetic heterogeneity in coastal communities implies increased population size, extending a previously proposed[10] urbanization model for the Late Viking Age city of Sigtuna (which suggested that more-cosmopolitan trading centres were already present at the end of the Viking Age in Northern Europe) both spatially and further back in time. The formation of large-scale trading and cultural networks that spread people, goods and warfare took time to affect the heartlands of Scandinavia, which retained pre-existing genetic differences into the Medieval period.

Finally, our findings show that Vikings were not simply a direct continuation of Scandinavian Iron Age groups. Instead, we observe gene flow from the south and east into Scandinavia, starting in the Iron Age and continuing throughout the duration of the Viking Age, from an increasing number of sources. Many Viking Age individuals—both within and outside Scandinavia—have high levels of non-Scandinavian ancestry, which suggests ongoing gene flow across Europe.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2688-8.

1. Brink, S. & Price, N. (eds) *The Viking World* (Routledge, 2008).
2. Jesch, J. *The Viking Diaspora* (Routledge, 2015).
3. Eriksen, M. H., Pedersen, U., Rundberget, B. & Axelsen, I. *Viking Worlds: Things, Spaces and Movement* (Oxbow Books, 2014).
4. Sindbæk, S. M. & Trakadas, A. *The World in the Viking Age* (Viking Ship Museum in Roskilde, 2014).
5. Sikora, M. et al. The population history of northeastern Siberia since the Pleistocene. *Nature* **570**, 182–188 (2019).
6. Lamnidis, T. C. et al. Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nat. Commun*. **9**, 5018 (2018).
7. Saag, L. et al. The arrival of Siberian ancestry connecting the eastern Baltic to Uralic speakers further east. *Curr. Biol*. **29**, 1701–1711.e16 (2019).
8. Hedeager, L. in *The Viking World* (eds Brink, S. & Price, N.) 35–46 (Routledge, 2008).
9. Hedeager, L. *Iron Age Myth and Materiality: An Archaeology of Scandinavia AD 400–1000* (Routledge, 2011).
10. Krzewińska, M. et al. Genomic and strontium isotope variation reveal immigration patterns in a Viking Age town. *Curr. Biol*. **28**, 2730–2738.e10 (2018).
11. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet*. **8**, e1002453 (2012).
12. Shepard, D. A Two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM National Conference* (eds Blue, R. B. & Rosenberg, A. M.) 517–524 (ACM, 1968).
13. Hansen, U. L. *Römischer Import im Norden: Warenaustausch zwischen dem Römischen Reich und dem freien Germanien während der Kaiserzeit unter besonderer Berücksichtigung Nordeuropas* (Det Kongelige Nordiske Oldskriftselskab, 1987).
14. Andersson, K. *I Skuggan av Rom: Romersk Kulturpåverkan i Norden* (Atlantis, (2013).
15. Bill, J. in *The Viking World* (eds Brink, S. & Price, N.) 170–180 (Routledge, 2008).
16. Sindbæk, S. M. The small world of the Vikings: networks in early Medieval communication and exchange. *Norw. Archaeol. Rev*. **40**, 59–74 (2007).
17. Hilberg, V. & Kalmring, S. in *Viking Archaeology in Iceland: Mosfell Archaeological Project* (eds Zori, D. & Byock, J.) 221–245 (Brepols, 2014).
18. Ebenesersdóttir, S. S. et al. Ancient genomes from Iceland reveal the making of a human population. *Science* **360**, 1028–1032 (2018).
19. Helgason, A. et al. mtDna and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am. J. Hum. Genet*. **68**, 723–737 (2001).
20. Downham, C. Viking ethnicities: a historiographic overview. *Hist. Compass* **10**, 1–12 (2012).
21. Fellows-Jensen, G. in *The Viking World* (eds Brink, S. & Price, N.) 391–400 (Routledge, 2008).
22. Bowden, G. R. et al. Excavating past population structures by surname-based sampling: the genetic legacy of the Vikings in northwest England. *Mol. Biol. Evol*. **25**, 301–309 (2008).
23. Leslie, S. et al. The fine-scale genetic structure of the British population. *Nature* **519**, 309–314 (2015).
24. Athanasiadis, G. et al. Nationwide genomic study in Denmark reveals remarkable population homogeneity. *Genetics* **204**, 711–722 (2016).
25. Loe, L., Boyle, A., Webb, H. & Score, D. *'Given to the Ground': A Viking Age Mass Grave on Ridgeway Hill, Weymouth* (Dorset Natural History and Archaeological Society, 2014).
26. Wallis, S. *The Oxford Henge and Late Saxon Massacre: With Medieval and Later Occupation at St John's College, Oxford* (Thames Valley Archaeological Services Limited, 2014).
27. Douglas Price, T., Frei, K. M., Dobat, A. S., Lynnerup, N. & Bennike, P. Who was in Harold Bluetooth's army? Strontium isotope investigation of the cemetery at the Viking Age fortress at Trelleborg, Denmark. *Antiquity* **85**, 476–489 (2011).
28. Price, T. D. & Arneborg, J. The peopling of the North Atlantic: isotopic results from Greenland. *Journal of the North Atlantic* **7**, 164–185 (2014).
29. Arneborg, J. in *The Viking World* (eds Brink, S. & Price, N.) 588–603 (Routledge, 2008).
30. Dugmore, A. J. et al. Cultural adaptation, compounding vulnerabilities and conjunctures in Norse Greenland. *Proc. Natl Acad. Sci. USA* **109**, 3658–3663 (2012).
31. Arneborg, J. in *Medieval Archaeology in Scandinavia and Beyond: History, Trends and Tomorrow* (eds Kristiansen, M. S. et al.) 247–271 (Aarhus Universitetsforlag, 2015).
32. Lynnerup, N. *The Greenland Norse: A Biological–Anthropological Study* (Museum Tusculanum, 1998).
33. Sindbæk, S. M. in *Routledge Handbook of Archaeology and Globalization* (ed. Hodos, T.) 553–565 (Routledge, 2016).
34. Peets, J. et al. Research results of the Salme ship burials in 2011–2012. *Archaeological Fieldwork in Estonia* **2012**, 43–60 (2012).
35. Douglas Price, T., Peets, J., Allmäe, R., Maldre, L. & Oras, E. Isotopic provenancing of the Salme ship burials in pre-Viking Age Estonia. *Antiquity* **90**, 1022–1037 (2016).
36. Cheng, J. Y., Racimo, F. & Nielsen, R. Ohana: detecting selection in multiple populations by modelling ancestral admixture components. Preprint at https://doi.org/10.1101/546408 (2019).
37. Alves, J. M. et al. Parallel adaptation of rabbit populations to myxoma virus. *Science* **363**, 1319–1326 (2019).
38. Enattah, N. S. et al. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet*. **30**, 233–237 (2002).

39. Bersaglieri, T. et al. Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet*. **74**, 1111–1120 (2004).
40. Allentoft, M. E. et al. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
41. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
42. Fearon, E. R. et al. Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* **247**, 49–56 (1990).
43. Siddiqa, A. et al. Regulation of CD40 and CD40 ligand by the AT-hook transcription factor AKNA. *Nature* **410**, 383–387 (2001).
44. Pedersen, C. B. et al. The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
45. Watanabe, K. et al. A global view of pleiotropy and genetic architecture in complex traits. *Nat. Genet*. **51**, 1339–1348 (2019).
46. Westerdahl, C. The maritime cultural landscape. *Int. J. Naut. Archaeol*. **21**, 5–14 (1992).
47. Hyenstrand, Å. *Ancient Monuments and Prehistoric Society* (Central Board of National Antiquities, 1979).
48. Callmer, J. in *Regions and Reflections: In honour of Märta Strömberg* (eds Jennbert, K. et al.) 257–273 (1991).
49. Jakobsen, J. G. G. & Dam, P. *Atlas Over Danmark: Historisk-Geografisk Atlas* (Det Kongelige Danske Geografiske Selskab, 2008).

---

[1]Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark. [2]Institute of Molecular Biology, National Academy of Sciences, Yerevan, Armenia. [3]Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark. [4]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [5]School of Statistical Sciences, University of Bristol, Bristol, UK. [6]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [7]The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [8]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland. [9]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [10]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [11]Institute of Biological Psychiatry, Mental Health Services Copenhagen, Copenhagen, Denmark. [12]HSE University, Russian Federation National Research University Higher School of Economics, Moscow, Russia. [13]Department of Archaeology and Ancient History, Lund University, Lund, Sweden. [14]Sydsvensk Arkeologi AB, Kristianstad, Sweden. [15]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. [16]Department of Genetics, University of Cambridge, Cambridge, UK. [17]Department of Health Technology, Section for Bioinformatics, Technical University of Denmark (DTU), Copenhagen, Denmark. [18]Department of Human Genetics, McGill University, Montréal, Quebec, Canada. [19]National Institute of Genomic Medicine (INMEGEN), Mexico City, Mexico. [20]Archaeological Research Collection, Tallinn University, Tallinn, Estonia. [21]Museum and Institute of Zoology, Polish Academy of Sciences, Warsaw, Poland. [22]Anuchin Research Institute and Museum of Anthropology, Moscow State University, Moscow, Russia. [23]Manx National Heritage, Douglas, Isle of Man. [24]Upplandsmuseet, Uppsala, Sweden. [25]NTNU University Museum, Department of Archaeology and Cultural History, Trondheim, Norway. [26]The Archaeologists, National Historical Museums, Stockholm, Sweden. [27]Thames Valley Archaeological Services (TVAS), Reading, UK. [28]Department of Archaeology and Ancient History, Uppsala University, Uppsala, Sweden. [29]Institute of Archaeology, University of Warsaw, Warsaw, Poland. [30]Department of Cultural Heritage Conservation, National Yunlin University of Science and Technology, Douliou, Taiwan. [31]Institute of Archaeology, Reykjavík, Iceland. [32]UHI Archaeology Institute, University of the Highlands and Islands, Kirkwall, UK. [33]Department of Bioarchaeology, Institute of Archaeology of National Academy of Sciences of Ukraine, Kiev, Ukraine. [34]Soprintendenza Archeologia, Belle Arti e Paesaggio per le Province di Barletta, Andria, Trani e Foggia, Foggia, Italy. [35]Jönköping County Museum, Jönköping, Sweden. [36]Trinity College Dublin, Dublin, Ireland. [37]Heritage Burial Services, Oxford Archaeology, Oxford, UK. [38]National Museum of Ireland, Dublin, Ireland. [39]Institute of Archaeology, Maria Curie-Sklodowska University in Lublin, Lublin, Poland. [40]Västergötlands Museum, Skara, Sweden. [41]Department of History and Archaeology, Amgueddfa Cymru–National Museum Wales, Cardiff, UK. [42]Trzy Epoki Archaeological Service, Klimontów, Poland. [43]Historical Faculty, Moscow State University, Moscow, Russia. [44]Museum of Southwest Jutland, Ribe, Denmark. [45]Department of Slavic–Finnish Archaeology, Institute for the History of Material Culture, Russian Academy of Sciences, Saint Petersburg, Russia. [46]National Museum of Denmark, Copenhagen, Denmark. [47]Department of Research and Heritage, Roskilde Museum, Roskilde, Denmark. [48]Langelands Museum, Langeland, Denmark. [49]Department of Humanities, University of Foggia, Foggia, Italy. [50]Department of Molecular Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway. [51]Department of Archaeology and Ancient History, Uppsala University Campus Gotland, Visby, Sweden. [52]Tjóðsavnið – Faroe Islands National Museum, Tórshavn, Faroe Islands. [53]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera), Russian Academy of Science, St Petersburg, Russia. [54]Malmö Museum, Malmö, Sweden. [55]Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Université de Toulouse, Université Paul Sabatier, Toulouse, France. [56]Department of Archaeology, Simon Fraser University, Burnaby, British Colombia, Canada. [57]Department of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark. [58]School of GeoSciences, University of Edinburgh, Edinburgh, UK. [59]Department of Natural History, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [60]Trace and Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin University, Perth, Western Australia, Australia. [61]Museum of Cultural History, University of Oslo, Oslo, Norway. [62]Centre for Urban Network Evolutions (UrbNet), School of Culture and Society, Aarhus University, Højbjerg, Denmark. [63]Institute of Archaeology, Conservation and History, Oslo, Norway. [64]Department of Historical Studies, University of Gothenburg, Gothenburg, Sweden. [65]Department of Integrative Biology, UC Berkeley, Berkeley, CA, USA. [66]Department of Statistics, UC Berkeley, Berkeley, CA, USA. [67]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Copenhagen, Denmark. [68]Department of Zoology, University of Cambridge, Cambridge, UK. [69]The Danish Institute for Advanced Study, University of Southern Denmark, Odense, Denmark. [70]The Wellcome Trust Sanger Institute, Cambridge, UK. [71]These authors contributed equally: Ashot Margaryan, Daniel J. Lawson, Martin Sikora, Fernando Racimo. ✉e-mail: rasmus_nielsen@berkeley.edu; Thomas.Werge@regionh.dk; ew482@cam.ac.uk

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Laboratory work

Laboratory work was conducted in the dedicated ancient DNA clean-room facilities at the Globe Institute (University of Copenhagen), according to strict ancient DNA standards[50,51]. The overwhelming majority of ancient samples were petrous bones and teeth (Supplementary Table 1). The details of DNA extraction can be found in Supplementary Note 2. Double-stranded blunt-end DNA libraries were prepared using Illumina-specific adapters and NEBNext DNA Sample Pre Master Mix Set 2 (E6070) kit. We used an Agilent Bioanalyzer 2100 to quantify the amount of the purified DNA libraries. The libraries were sequenced 80-bp single-read chemistry on Illumina HiSeq 2500 machines at the Danish National High-throughput DNA Sequencing Centre.

Bioinformatics analysis and quality assessment

We used AdapterRemoval v.2.1.3[52] for removing Illumina adaptor sequences, keeping only sequences with a minimum length of 30 bp. Adaptor-free sequences were mapped against the human reference genome build 37 using BWA v.0.7.10 aligner[53] with the seed (-l parameter) disabled for higher sensitivity of ancient DNA reads[54]. DNA sequences were processed with samtools v.1.3.1[53], and only sequences with mapping quality ≥ 30 were kept. Picard v.1.127 (http://broadinstitute.github.io/picard) was used to sort the reads and remove duplicates. DNA libraries were combined at the sample level and realigned using GATK v.3.3.0[55] with Mills and 1000G gold-standard insertions and deletions (indels). At the end, realigned .bam files had the md-tag updated and extended base alignment qualities calculated using samtools calmd. Read depth and coverage were determined using pysam (http://code.google.com/p/pysam/) and BEDtools[56]. The mapping statistics for the ancient samples are summarized in Supplementary Table 2.

We used mapDamage v.2.0 to obtain approximate Bayesian estimates of damage parameters[57]. Data authenticity was assessed by estimating the rate of mismatches to the consensus mitochondrial sequence using contamMix[58] and Schmutzi[59], as well as the excess of heterozygous positions in male haploid X chromosomes using ANGSD[60]. The sex of ancient individuals was determined by calculating the R_γ parameter[61].

Uniparental haplogroup determination and kinship analysis

The mitochondrial haplogroups of the ancient individuals were assigned using haplogrep[62]. To get the mtDNA consensus sequences, we aligned the trimmed reads of ancient samples to the human mitochondrial reference genome: revised Cambridge Reference Genome (rCRS). Base quality ≥ 20 and mapping quality ≥ 30 filtering options were applied. Only SNPs at sites ≥3× coverage were considered for consensus calling using samtools mpileup/bcftools v.1.3.1[53].

We identified male Y chromosome lineages using the pathPhynder workflow (https://github.com/ruidlpm/pathPhynder) and Yleaf v.2[63]. For the latter, the analysis was restricted to 26,083 biallelic SNPs from the International Society of Genetic Genealogy (ISOGG) 2019 database (https://isogg.org/tree/ISOGG_YDNA_SNP_Index.html).

We used NgsRelate[64] to detect family relationships between all pairs of individuals. NgsRelate is a maximum-likelihood based program that—for a pair of individuals based on genotype likelihoods—estimates the three coefficients, $k0$, $k1$ and $k2$, which denote the proportions of the genome in which the pair of analysed individuals share 0, 1 and 2 alleles identical-by-descent, respectively. We only included the 376 samples with sequencing depth above 0.1× for the analysis. From these, we estimated genotype likelihoods and allele frequencies with ANGSD[60] using the SAMtools genotype likelihood model (-gl 1) including reads with mapping quality ≥ 30 and bases with base quality ≥ 20. We estimated genotype likelihoods and allele frequencies only for the autosomal transversion sites for which the 1000 Genomes CEU population (Utah residents with northern and western European ancestry) has a minor allele frequency of 0.05, resulting in 1,752,719 sites. READ[65] was used to confirm the degree of relatedness between pairs of individuals. The pedigree reconstructions on the basis of the kinship coefficients were conducted using Pedigree Reconstruction and Identification of a Maximum Unrelated Set (PRIMUS)[66].

Imputation

We imputed the genotypes of 298 ancient samples (289 from this study, and 9 from a previous study[10]) that had a sequencing depth greater than 0.5×. We used Beagle v.4.1[67] for imputations based on the genotype likelihood data, which was first estimated by GATK v.3.7.0 UnifiedGenotyper. To generate the genotype data, we called only biallelic sites present in the 1000 Genomes dataset, and only the observed alleles (--genotyping_mode GENOTYPE_GIVEN_ALLELES). The resulting .vcf files were filtered by setting genotype likelihoods to 0 for all three genotypes (for example, hom ref, het and hom alt) for sites with potential deamination (C>T and G>A), as described in a previous study[68]. Following this, the per-individual .vcf files were merged using bcftools v.1.3.1. The combined .vcf files were then split into 15,000 markers each and imputed separately using Beagle 4.0 using the 1000 Genomes phase-3 map included with Beagle (*.phase3.v5a.snps.vcf.gz and plink.chr*.GRCh37.map) with input through the genotype likelihood option. Run time for imputing using Beagle was approximately 280,000 core hours.

Merge with existing panels

Scandinavian panel. To assess the genetic relationships of various Viking Age groups with their present-day counterparts, we constructed a reference panel enriched with Scandinavian populations on the basis of published datasets: the EGAD00010000632 data set from a previous publicaton[23] (UK dataset) and the EGAD00000000120 dataset from The International Multiple Sclerosis Genetics Consortium and The Wellcome Trust Case Control Consortium 2 (ref. [69]) (EU dataset) (see Supplementary Note 6 for details). The seven most relevant populations from Denmark, Sweden, Norway, Finland, Poland, UK and Italy were considered ($n = 1{,}464$) with a total number of 414,264 SNPs. The Han Chinese (CHB) and Yoruba (YRI) populations from the 1000 Genomes project phase-3 database were merged to this panel as outgroups.

The 1000 Genomes panel. We used a set of 1,995 individuals from 20 populations (excluding individuals from the AMR super-population, as well as admixed ASW and ACB populations) of the 1000 Genomes project phase-3 release 5 (ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/). We restricted the dataset to a set of 12,731,663 biallelic transversion SNPs located within the strict mappability mask regions (ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/accessible_genome_masks/).

Analyses of phenotype associated SNPs were carried out using five European-ancestry populations: Spanish (IBS), Tuscan (TSI), CEU, British (GBR) and Finnish (FIN), along with CHB and YRI as outliers. These were used to assess genome-wide allele frequencies for various SNPs associated with pigmentation phenotypes and lactose intolerance.

Ancient panels. We constructed datasets for population genetic analyses by merging the newly sequenced Viking Age individuals as well as other previously published ancient individuals[40,41,68,70–96] with the two modern reference panels. Ancient individuals were represented with pseudohaploid genotypes, by using mpileup command of samtools and randomly sampling an allele passing filters (mapping quality ≥ 30 and base quality ≥ 30), further requiring that it matched one of the two alleles observed in the reference panel (Supplementary Table 3).

Clustering analyses

On the basis of the pseudohaploid individuals from the ancient panels, we ran ADMIXTURE[97] by thinning the dataset for linkage disequilibrium using plink with recommended settings (--indep-pairwise 50 10 0.1). This dataset contained 1,324 individuals for 151,235 markers for the autosomal chromosomes. Only samples with >20,000 SNPs overlapping with the Human Origins panel were kept in the analysis, resulting in 378 samples from this study. We did 50 replicates with different seeds for $k = 2$ to $k = 10$. We used pong[98] to identify the best run for each k and similar components between different k values.

The large number of ancient individuals included in the analysis panels facilitates genetic clustering using the ancient individuals themselves, rather than projecting them on axes of variation inferred from modern populations. We carried this out using MDS on a distance matrix obtained from pairwise identity-by-state sharing between individuals, using the cmdscale function in R. We performed the main genetic clustering on a set of 1,306 ancient Eurasian individuals with >50,000 SNPs with genotype data, restricting to the batch-corrected SNP set described in Supplementary Note 8. Results from the batch-corrected MDS were combined with further dimensionality reduction using UMAP, implemented in the uwot package in R.

Population genetics

We used f_4 statistics to investigate allele-sharing between sets of test individuals and different modern and ancient groups (Supplementary Note 9). To characterize the deep ancestry relationship of the study individuals we calculated f_4(YRI, test individual; Barcin_EN.SG, Yamnaya_EBA.SG) for all ancient Europeans from the Bronze Age onwards (1000 Genomes panel merge). This statistic contrasts genetic affinities of the test individuals with two major ancestry groups that contributed to the gene pool of ancient Europeans from the Bronze Age onwards: Anatolian farmers and Steppe pastoralists. Genetic continuity with Scandinavian Iron Age groups was investigated using f_4(YRI, test group; test individual, Scandinavia Iron Age group) (1000 Genomes panel merge). This statistic measures whether a test individual is consistent with forming a clade with Scandinavian Iron Age groups to the exclusion of a test group from outside of Scandinavia. Genetic affinities between ancient groups and present-day populations were investigated using f_4(YRI, test individual; present-day test population, present-day reference population) (Scandinavian panel).

Ancestry modelling using qpAdm

We estimated ancestry proportions of Viking Age groups using qpAdm[70], which is based on f_4-statistics of the from $f_4(X,O1;O2,O3)$, in which X is either the source or target population, and $O1$, $O2$ and $O3$ are triplets of outgroups to the source and target groups. To minimize batch effects and/or biases due to ancient DNA damage or SNP ascertainment, we used a set of 1,800,038 transversion-only sites that were found polymorphic with minor allele frequency ≥ 0.5% and missing genotype rate of ≤ 15% in the 1000 Genomes panel merge.

Genetic diversity

The genetic diversity of ancient groups was assessed using conditional nucleotide diversity, as previously described[73]. For this analysis, pairwise differences between individuals were calculated using SNPs polymorphic in an outgroup population (YRI) and with a minor allele count ≥ 5 in the 1000 Genomes merge.

Identity-by-descent analysis

The imputed genotypes of 298 individuals were used to infer genomic segments shared via identity-by-descent within the context of a reference panel of 1,464 present-day Europeans, using IBDseq[99] (version r1206) with default parameters. We conducted genetic clustering by MDS on a distance matrix obtained from pairwise identity-by-descent sharing and UMAP to reveal fine-scale population structure among Viking Age individuals.

Painting

To assess the fine-scale variation in genetic ancestry proportions of Viking Age individuals we used Chromosome Painting[11]. The following describes the general workflow of the Chromosome Painting analysis (see Supplementary Note 11 for details).

First, we created a modern reference panel using 1,675 modern individuals sampled from northern Europe, using the standard FineSTRUCTURE pipeline. We applied ChromoPainter to paint all modern individuals using the remaining individuals as donors using fs2.0.8. Related individuals were identified through increased haplotype similarity, and admixed individuals were identified by their FineSTRUCTURE clustering. These were removed, leading to 1,554 unrelated individuals who were re-painted. These individuals were then clustered using FineSTRUCTURE, resulting in 40 populations. After removal of small populations and merging of the CHB and YRI subpopulations, this resulted in 23 modern populations with geographical meaning. We named the resulting clustering the modern reference panel, which consists of 23 modern surrogate populations and 23 modern donor populations (Supplementary Fig. 11.2).

Second, we created an ancient reference panel using the modern reference panel, by applying ChromoPainter to paint all ancient individuals using the modern population palette (Supplementary Fig. 11.3). We then created a supervised ancient population palette consisting of 14 populations which either (a) represent a modern ancestry direction or (b) are best associated with a modern ancestry direction. The paintings consider the average per-individual donor rate to each of the seven modern populations, normalizing each donor label to have a mean of 1 (Supplementary Fig. 11.4). The individuals that contribute most to a population represent it (above a threshold amount chosen by identifying a change point). The remaining individuals are assigned to the population that they are best associated with. We create an ancient population surrogate for each modern population, consisting of the individuals that represent each modern population. For $k = 7$ modern populations, this results in a matrix of $k = 7$ rows (surrogate populations) and $2K = 14$ columns (donor palette populations), which captures the ancient population structure (Supplementary Fig. 11.6).

Third, we inferred ancestry by learning about population structure in modern individuals or ancient individuals, painting them with respect to the ancient population panel and fitting them as a mixture using the ancient population surrogates, using the non-negative least squares implemented in GLOBETROTTER[100] (Supplementary Information section 11) with uncertainty estimated using 100 bootstrap replicates. All samples were analysed by leaving out one individual per donor population so that modern and ancient individuals are exchangeable (as the ancient individual is itself excluded from its own ancient donor population). We report this in a number of ways. The inferred ancestry results (Supplementary Table 6) are summarized by taking the mean across inferred populations in Supplementary Fig. 11.11; Supplementary Fig. 11.12 shows the means over sample information labels. We performed a spatiotemporal regression (Supplementary Table 11.2) using the model $a_{ik} = \alpha_{jk} t_i + \beta_{jk} x_i + \gamma_{jk} y_i + \varepsilon_{ijk}$ in which a_{ik} is the amount of ancestry individual i possesses from population k in regional analysis j, t_i is the age category of the individual (1 = Iron Age, 2 = Early Viking Age, 3 = Viking Age, 4 = Medieval) and x_i and y_i are the longitude and latitude of the burial location of the individual, respectively. The modern ancestry results are estimated using the spatial median instead of the mean, to account for ancestry being constrained in a k-dimensional simplex (Supplementary Fig. 11.14), with uncertainty quantified by bootstrap resampling of individuals (Supplementary Fig. 11.15).

Fourthly, we performed sensitivity analyses to ensure that the inference procedure performed as expected. We checked that sequence depth was not associated with cluster membership (Supplementary

Fig. 11.7), and that sequence depth did not significantly affect inferred ancestry (Supplementary Fig. 11.8) by downsampling individuals with high-depth data available, rephasing, re-imputing and repainting them, and assigning ancestry using the above procedure. Results 2× and above were extremely similar, whereas at 1× there was some loss of precision but the broad structure remained clear.

Finally, we ran a principal components analysis of the ancient + modern populations painted against our donor populations (Supplementary Fig. 11.9) as well as an all-versus-all ChromoPainter analysis including modern and ancient individuals (Supplementary Fig. 11.10).

Ancestry diversity measure

We wish to quantify diversity in ancestry for a population of individuals, with diverse meaning a large deviation of individual ancestry estimates from the average ancestry in that population. We compute the average Kullback–Leibler divergence for each individual label from the average of that label:

$$D(A^{(l)}) = \frac{1}{n_l}\sum_{i=1}^{n_l} KL(A_i^{(l)} \;||\; p^{(l)})$$

in which $A^{(l)}$ is the n_l by K matrix of ancestry estimates in label l, $p(l)$ is the length **K** vector of average ancestries in that label, and $KL(Q\,||\,P) = \sum_{k=1}^{K} q_k \log_2\left(\frac{q_k}{p_k}\right)$. We performed a simulation study to validate this measure (Supplementary Information section 11, Supplementary Fig. 11.13), which allowed us to calibrate the expected diversity as a function of sample size.

Spatiotemporal patterns

To visualize the migration patterns of the Vikings, we used inverse distance weighting interpolation—implemented in the function idw of the R package gstat—to interpolate the proportion of each ancient genome that was attributed by our fineSTRUCTURE analysis (Supplementary Table 6) to one of the predefined ancestry groups: UK, Denmark, Norway, Sweden, Italy, Poland and Finland. We used the Shepard method of interpolation[12,101] with the weight for a given interpolation location x equal to $1/(d(x,\upsilon)^2)$, in which υ is the location of an observed sample and $d(a,b)$ is the distance between two points a and b. For plotting maps, we used a Mercator projection and downloaded coastal contours at 1:50-m scale from Natural Earth (https://www.naturalearthdata.com/).

Lactase persistence and pigmentation SNPs

For ancient populations we estimated the derived A allele frequency of the SNP rs4988235, known to affect expression of the lactase (*LCT*) gene. The ancestral G allele is responsible for lactase intolerance in adult Europeans[39]. We used ANGSD[60] to estimate the allele frequencies of the ancient population on the basis of the genotype likelihood data. We used the five European populations (CEU, FIN, GBR, TSI and IBS) and two outgroups (YRI and CHB) from the 1000 Genomes Project as comparative groups. We also included the present-day Danish population from the IPSYCH case–cohort study[44] and geographically proximate Iron and Bronze Age populations to trace frequency shifts of SNP rs4988235 through time. We also used ANGSD[60] to estimate the frequencies of 22 SNPs (HIrisPlex[102]) with strongest influence on human pigmentation phenotypes in the Viking Age and Early Viking Age Scandinavian population.

Signatures of selection

We aimed to find SNPs with allele frequencies that changed significantly in the last 10,000 years, using our ancient human genomes to look at the frequencies of alleles in the past. We combined our Viking Age and Iron Age genomes with previously published present-day, Bronze Age, Neolithic and Mesolithic sequence data typed at the Human Origins array (Supplementary Note 6). We filtered for genomes that were younger than 8000 BC and that were located within a bounding box encompassing the European continent: 30°< latitude < 75° and −15°< longitude < 45°. We then used neoscan in Ohana[36,103] to scan for variants with allele frequencies that were strongly associated with time, after controlling for genome-wide changes in ancestry that might have also occurred over time. We analysed only sites with a minor allele frequency >1% (Supplementary Note 14).

Tracking the evolution of complex traits in Scandinavia

We wanted to examine whether we could identify signals of recent population differentiation of complex traits by comparing genotypes of Viking Age samples excavated in Scandinavia (that is, Denmark, Sweden and Norway) with those of a present-day Scandinavian population. For the latter, we used imputed genotypes from subjects born in Denmark between 1981 and 2011 from the IPSYCH case–cohort study[44]. We downloaded summary statistics from the genome-wide association study ATLAS webpage (https://atlas.ctglab.nl)[45], from studies of 16 disease- and anthropometric traits (excluding those related to cognition) published in 2017 or later with SNP heritability estimated at >0.1, sample size of >100,000 and >100 identified genome-wide significant loci. We calculated polygenic risk scores based on independent (R^2 < 0.1 within 10-Mb range) genome-wide significant allelic effects and standardized them to a unit representing the standard deviation of the mean of their distribution. We then removed outliers (anyone with a value for any of the 25 principal components falling more than 4 standard deviations away from the group mean) reiteratively from within each ancestry group (treating the Scandinavian Viking age samples as one ancestry group), and subsequently tested for difference in polygenic risk score distribution between Viking Age samples and Danish-ancestry IPSYCH random population samples using a linear regression model correcting for sex and the 25 principal components.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequence data are available at the European Nucleotide Archive under accession number PRJEB37976.

Code availability

Functions for calculating *f*-statistics are available as an R package at GitHub (https://github.com/martinsikora/admixr).

50. Willerslev, E. & Cooper, A. Ancient DNA. *Proc. R. Soc. Lond. B* **272**, 3–16 (2005).
51. Gilbert, M. T. P., Bandelt, H.-J., Hofreiter, M. & Barnes, I. Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541–544 (2005).
52. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
54. Schubert, M. et al. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
55. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
56. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
57. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
58. Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
59. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
60. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
61. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).
62. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).

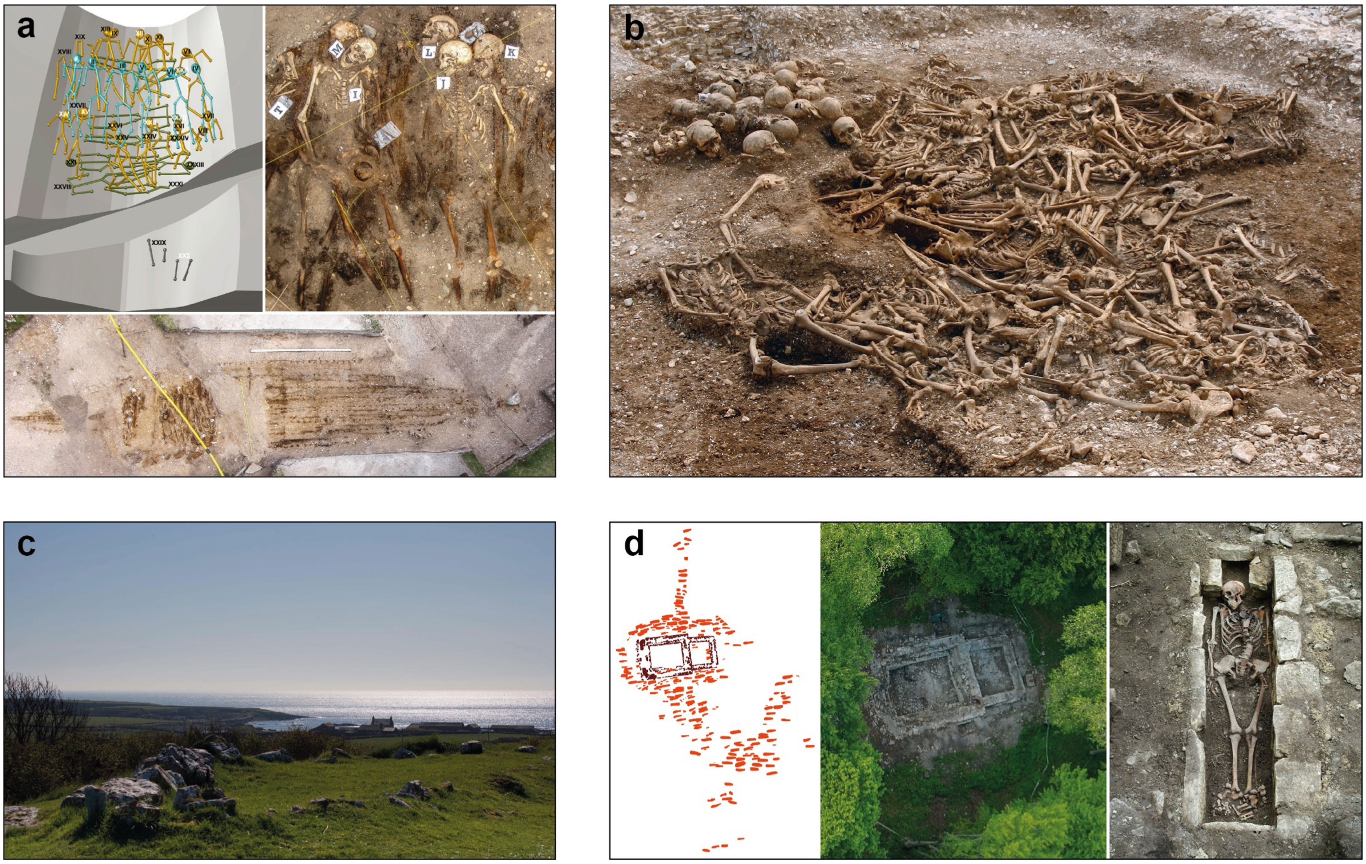
63. Ralf, A., Montiel González, D., Zhong, K. & Kayser, M. Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data. *Mol. Biol. Evol*. **35**, 1291–1294 (2018).
64. Korneliussen, T. S. & Moltke, I. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics* **31**, 4009–4011 (2015).
65. Monroy Kuhn, J. M., Jakobsson, M. & Günther, T. Estimating genetic kin relationships in prehistoric populations. *PLoS ONE* **13**, e0195491 (2018).
66. Staples, J., Nickerson, D. A. & Below, J. E. Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. *Genet. Epidemiol*. **37**, 136–141 (2013).
67. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet*. **81**, 1084–1097 (2007).
68. Martiniano, R. et al. The population genomics of archaeological transition in west Iberia: investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genet*. **13**, e1006852 (2017).
69. International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
70. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
71. Gamba, C. et al. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun*. **5**, 5257 (2014).
72. Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun*. **6**, 8912 (2015).
73. Skoglund, P. et al. Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014).
74. Schiffels, S. et al. Iron Age and Anglo-Saxon genomes from east England reveal British migration history. *Nat. Commun*. **7**, 10408 (2016).
75. Olalde, I. et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
76. Sikora, M. et al. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662 (2017).
77. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
78. Jones, E. R. et al. The Neolithic transition in the Baltic was not driven by admixture with early European farmers. *Curr. Biol*. **27**, 576–582 (2017).
79. Seguin-Orlando, A. et al. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
80. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
81. Hofmanová, Z. et al. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl Acad. Sci. USA* **113**, 6886–6891 (2016).
82. Damgaard, P. B. et al. 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).
83. Günther, T. et al. Population genomics of Mesolithic Scandinavia: investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol*. **16**, e2003703 (2018).
84. Mittnik, A. et al. The genetic prehistory of the Baltic Sea region. *Nat. Commun*. **9**, 442 (2018).
85. Kılınç, G. M. et al. The demographic development of the first farmers in Anatolia. *Curr. Biol*. **26**, 2659–2666 (2016).
86. Lazaridis, I. et al. Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
87. de Barros Damgaard, P. et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
88. Valdiosera, C. et al. Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proc. Natl Acad. Sci. USA* **115**, 3428–3433 (2018).
89. Martiniano, R. et al. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nat. Commun*. **7**, 10326 (2016).
90. Mathieson, I. et al. The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).
91. Gallego Llorente, M. et al. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**, 820–822 (2015).
92. Broushaki, F. et al. Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 499–503 (2016).
93. Veeramah, K. R. et al. Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. *Proc. Natl Acad. Sci. USA* **115**, 3494–3499 (2018).
94. Amorim, C. E. G. et al. Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat. Commun*. **9**, 3547 (2018).
95. Olalde, I. et al. A Common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol. Biol. Evol*. **32**, 3132–3142 (2015).
96. Olalde, I. et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
97. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. **19**, 1655–1664 (2009).
98. Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P. & Ramachandran, S. pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* **32**, 2817–2823 (2016).
99. Browning, B. L. & Browning, S. R. Detecting identity by descent and estimating genotype error rates in sequence data. *Am. J. Hum. Genet*. **93**, 840–851 (2013).
100. Hellenthal, G. et al. A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
101. Fortin, M.-J. & Dale, M. R. T. *Spatial Analysis: A Guide for Ecologists* (Cambridge Univ. Press, 2005).
102. Walsh, S. et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet*. **7**, 98–115 (2013).
103. Cheng, J. Y., Mailund, T. & Nielsen, R. Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics* **33**, 2148–2155 (2017).

Acknowledgements This work was supported by the Mærsk Foundation, the Lundbeck Foundation, the Novo Nordisk Foundation, the Danish National Research Foundation, University of Copenhagen (KU2016) and the Wellcome Trust (grant no. WT104125MA). E.W. thanks St John's College, Cambridge for providing an excellent environment for scientific thoughts and collaborations. S.R. was supported by the Novo Nordisk Foundation (NNF14CC0001). F.R. was supported by a Villum Fonden Young Investigator Award (project no. 00025300). G.S. and E.C. were supported by a Marie Skłodowska-Curie Individual Fellowship 'PALAEO-ENEO', a project funded by the European Union EU Framework Programme for Research and Innovation Horizon 2020 (grant agreement number 751349). R.M. was supported by an EMBO Long-Term Fellowship (ALTF 133-2017). M.C. is supported by the Canada Research Chairs Program (231256), the Canada Foundation for Innovation (36801) and the British Columbia Knowledge Development Fund (962-805808). I. Moltke was supported by a YDUN grant from Independent Research Fund Denmark (DFF-4090-00244) and a Villum Fonden Young Investigator Award (project no. 19114). N.P. and C.H.J. are supported by the Swedish Research Council (2015-00466). N.G. was supported by the Program of Fundamental Scientific Research of the State Academies of Sciences, Russian Federation, state assignment no. 0184-2019-0006. D.G.B. and L.M.C. were supported by Science Foundation Ireland/Health Research Board/Wellcome Trust award no. 205072. We thank the iPSYCH Initiative, funded by the Lundbeck Foundation (grant numbers R102-A9118 and R155-2014-1724), for supplying SNP frequency estimates from the present-day Danish population for comparison with Viking Age samples; M. Jakobsson and A. Götherström for providing preliminary access to the sequencing data of 23 Viking Age samples from Sigtuna; M. Corrente for providing access to the skeletal remains from Cancarro; and N. M. Mangialardi and M. Maruotti for the useful suggestions; Greenland National Museum and Archives, as well as the Gotland Museum, for permission to sample their skeletons; J. Kavanagh for providing information on his excavation, and L. Buckley, D. Keating and B. Ó Donnabháin for analysing the remains; R. Breward and J. Murden from the Dorset County Museum for allowing access to their assemblage for DNA sampling; J. Hansen and M. B. Henriksen at Odense Bys Museer for allowing sampling of skeletal material from Hessum and Galjedil; Moesgaard Museum for allowing sampling of skeletal material from Hesselbjerg; C. Bertilsson, P. Lingström, B. Lundberg, K. Lidén and J. Andersson for their help in sampling the ancient human remains; L. Drenzel for permission to sample the human remains; C. Ödman for suggesting relevant material for this study; Ł. Stanaszek, M. Zaitz and the Regional Museum in Cedynia for providing samples; L. Vinner, A. Seguin-Orlando, K. Magnussen, L. Petersen, C. Mortensen and M. J. Jacobsen at the Danish National Sequencing Centre for producing the analysed sequences; P. S. Olsen and T. Brand for technical assistance in the laboratories; R. M. Durbin and J. H. Barrett for comments and suggestions; and J. Wilson, J. Jesch, E. Harlitz-Kern and F. Martín Racimo for their feedback.

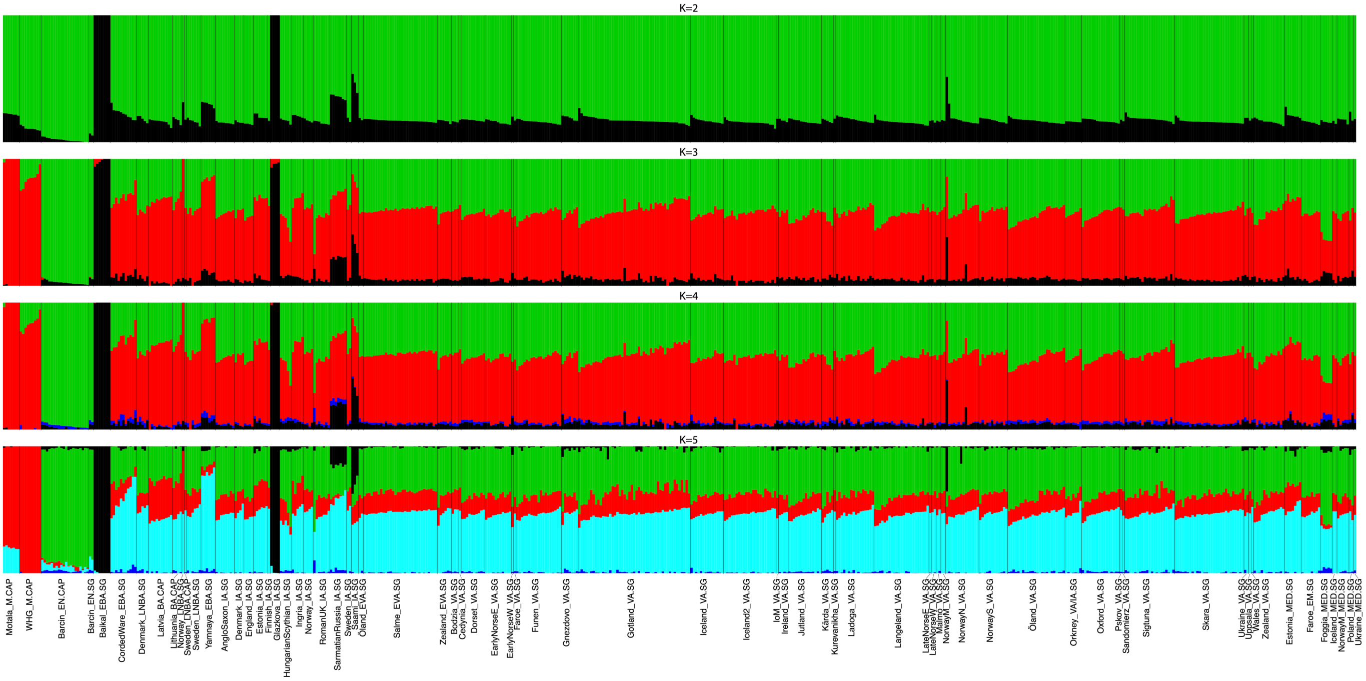
Author contributions E.W. initiated and led the study. E.W., A.M., D.J.L., Martin Sikora, F.R., R.N., K.K., L.H., S.M.S., J.B., N.P., T.W., A.I., M.E.A., M.W.P., N.L., J.A., I. Moltke and A.A. designed the study. A.M., P.d.B.D., L.M.C., M.M.B., A.K.F., I.L. and J.S. produced the data. A.M., D.J.L., Martin Sikora, F.R., S.R., I. Moltke, R.N., T.W., L.M.C., E.J., A.I., M.W.P., T.K., R.M., G.R., C.B., J.V.M.-M., H.M., A.A., J.C., K.H.I. and M.E.A. analysed or assisted in analysis of data. E.W., A.M., D.J.L., Martin Sikora, F.R., S.M.S., K.K., L.H., R.N., M.C. and A.I. interpreted the results with considerable input from I. Moltke, M.E.A., M.W.P., T.K., H.W., R.M., G.R., T.W., C.H.J., J.A., N.L., N.P., J.B., A.A., M.T.P.G., L.O. and other authors. E.W., A.M., D.J.L., Martin Sikora, F.R., S.M.S., K.K. and L.H. wrote the manuscript with considerable input from M.C., J.B., N.P., I. Moltke, N.L., A.I., R.M., E.J., J.A., M.L.J., C.H.J., M.W.P., M.E.A., G.R. and M.M., with contributions from all authors. A.M., L.M.C., M.W.P., H.W., M.M.B., P.d.B.D., A.K.F., M.A., R.A., M.M., E.C., G.S., A.B., A.F., B. Schütz, B. Skar, C.A., C.F., D.B., D.P., G.T.-W., H.G., I.L., I.G., I. Mainland, I.P., I.M.M., J.M., J. Gibson, J.P., J. Gustafsson, L. Simpson, L. Strand, L.L., Maeve Sikora, M.F., M.V., M.R., M.B., T.P., M. Søvso, N.G., T.C., O.K., O.U., P.F., P.H., S.S., S.V.A., S.E., V.M., W.B., Y.M., P.P., M.D.J., A.P., D.G.B., M.L.J., J.A., N.L., N.P., M.T.P.G., M.E.A., J.B. and E.W. excavated, curated, sampled and/or described analysed skeletons.

Competing interests The authors declare no competing interests.

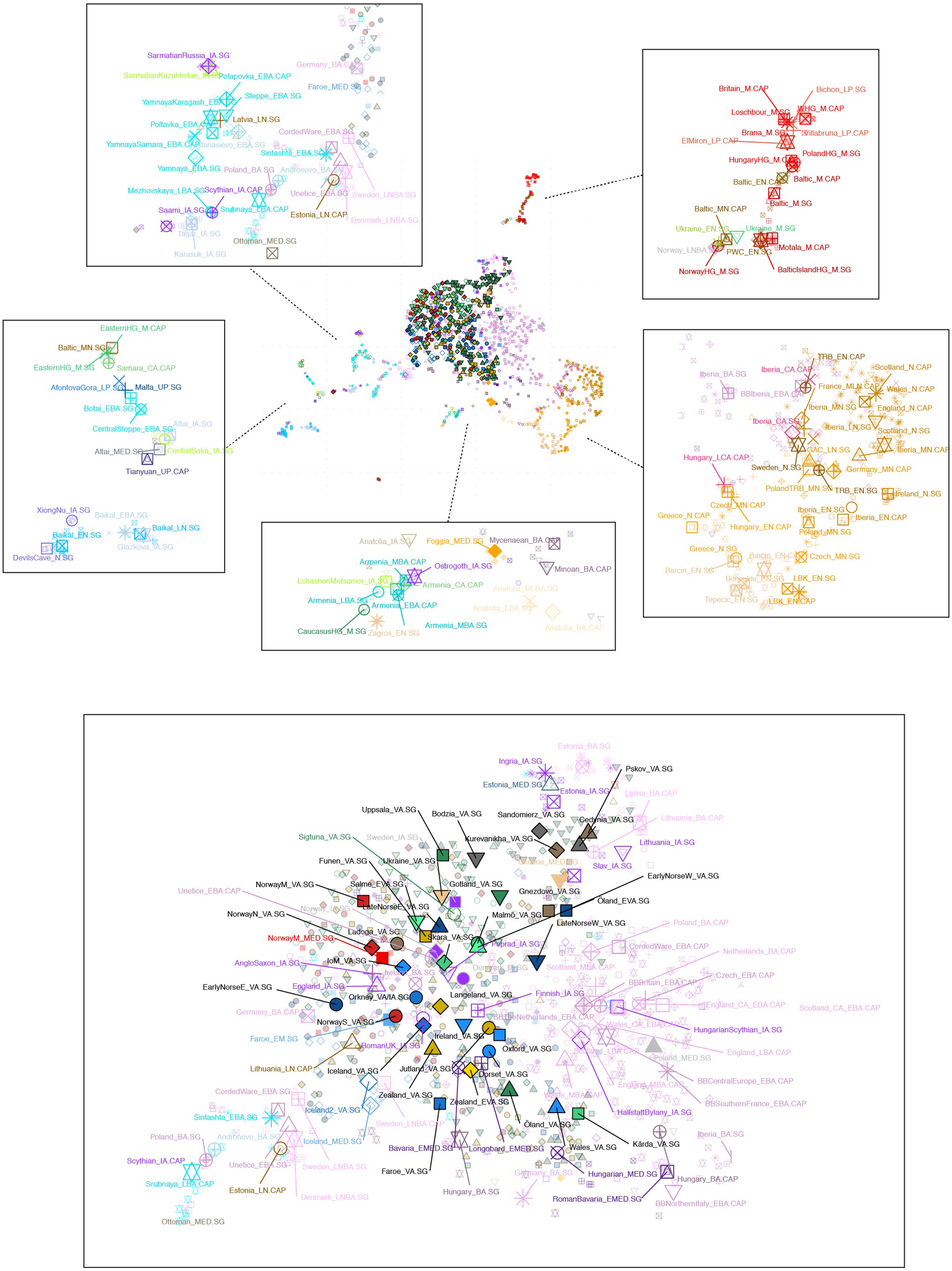
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2688-8.
Correspondence and requests for materials should be addressed to R.N., T.W. or E.W.
Peer review information *Nature* thanks James Barrett, Wolfgang Haak and Pontus Skoglund their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



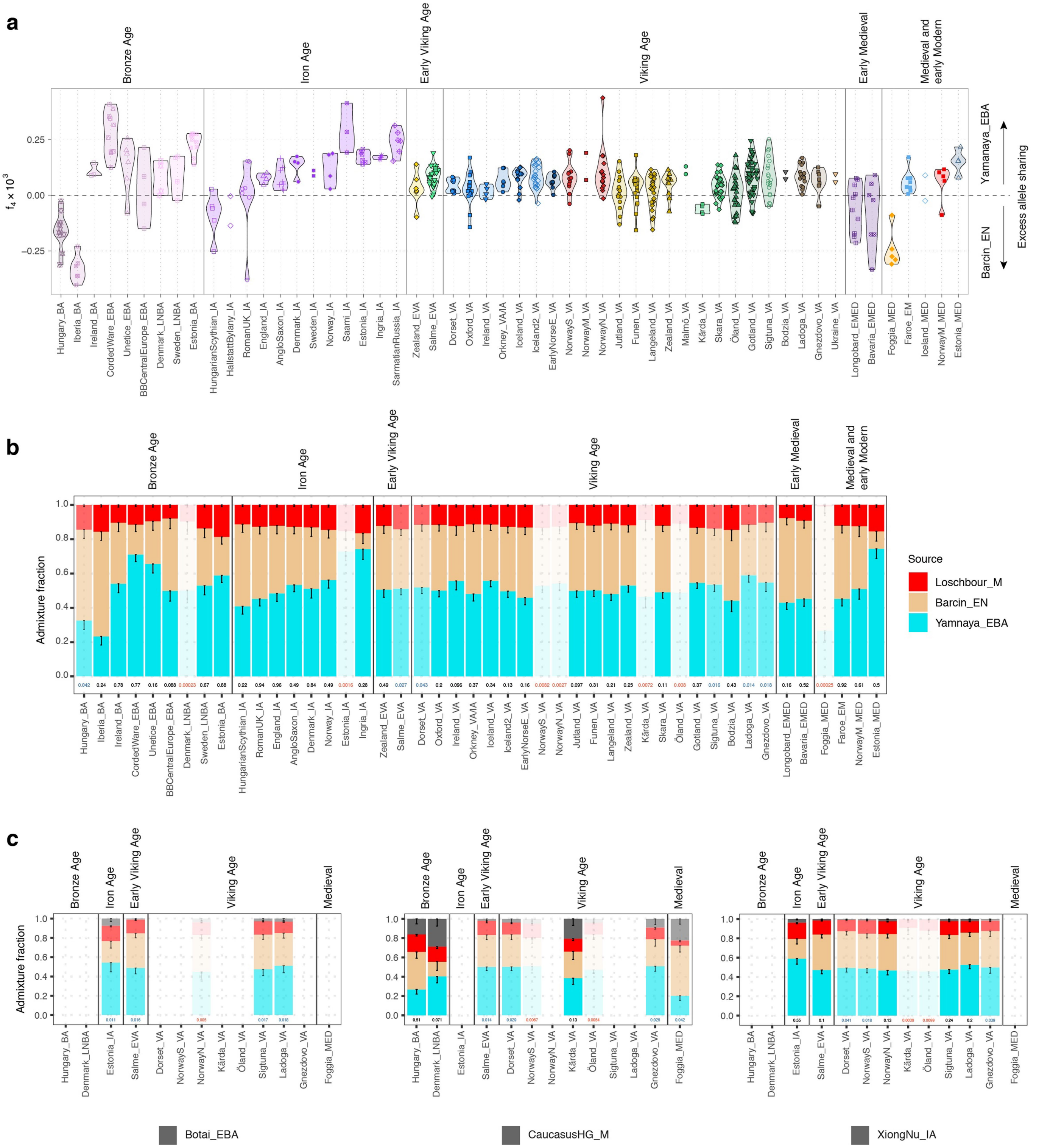
Extended Data Fig. 1 | Viking Age archaeological sites. Examples of a few archaeological Viking Age sites and samples used in this study. **a**, Salme II ship burial site of the Early Viking Age, excavated in present-day Estonia: schematic of skeletons (top left) and aerial images of skeletons (top right, and bottom). **b**, Ridgeway Hill mass grave dated to the tenth or eleventh century AD, located on the crest of Ridgeway Hill near Weymouth, on the south coast of England (reproduced with permission from Dorset County Council/Oxford Archaeology). Around 50 predominantly young adult male individuals were excavated. **c**, The site of Balladoole, around AD 900, a Viking was buried in an oak ship at Balladoole (Arbory) in the south east of the Isle of Man. **d**, Viking Age archaeological site in Varnhem, in Skara municipality (Sweden). Schematic map of the church foundation (left) and the excavated graves (red markings) at the early Christian cemetery in Varnhem; foundations of the Viking Age stone church in Varnhem (middle) and the remains of a 182-cm-long male individual (no. 17) buried in a lime stone coffin close to the church foundations (right).



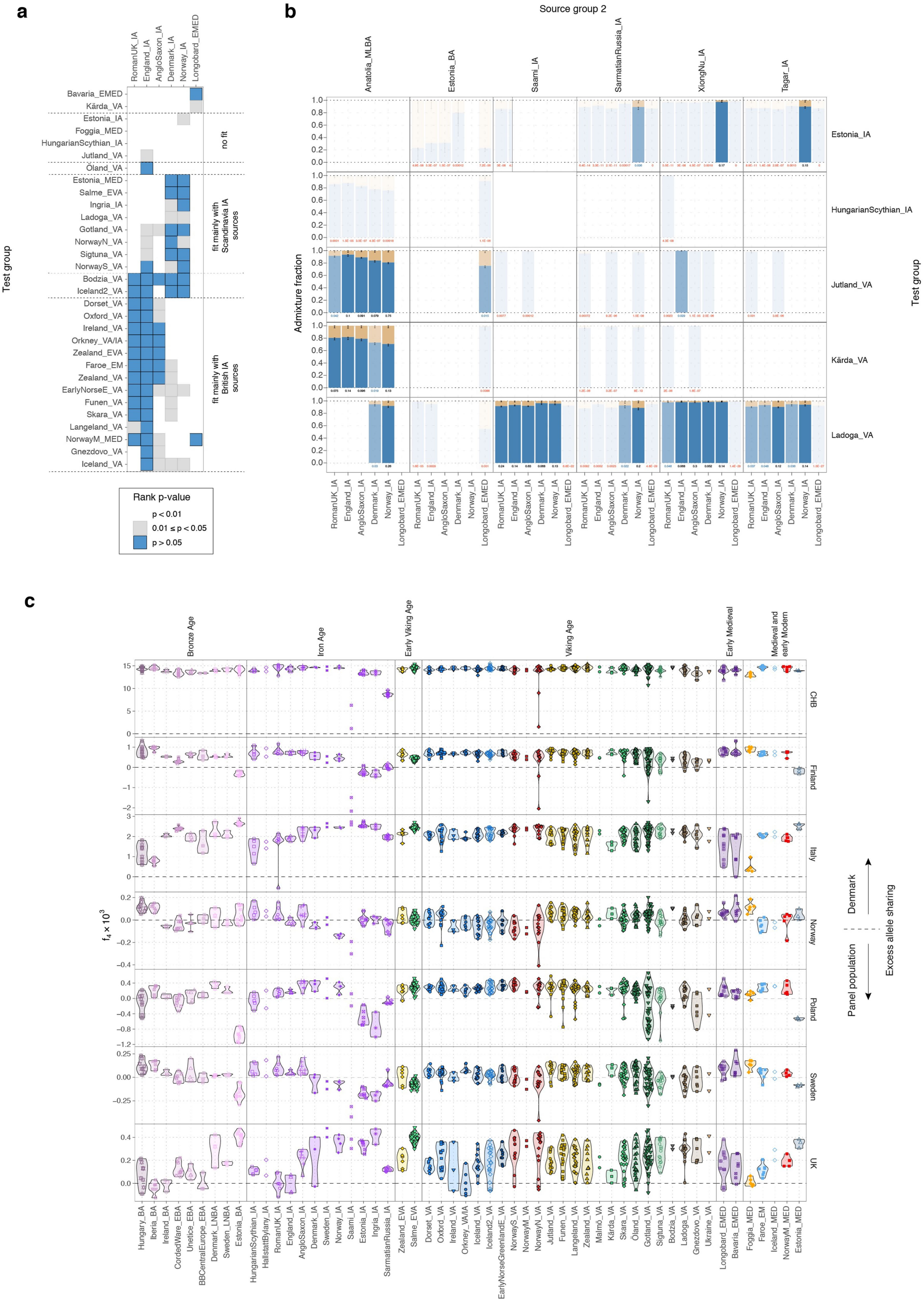
Extended Data Fig. 2 | Model-based clustering analysis. Admixture plot ($K=2$ to $K=5$) for 567 ancient individuals, spanning 71 populations. This figure is a subset of the most relevant individuals and populations from Supplementary Fig. 7.2; see Supplementary Note 7 for further details. This plot consists of 378 ancient samples from this study; Viking Age samples from Sigtuna (Sweden)[10] ($n=21$), Iceland[18] ($n=22$) and other ancient comparative groups ($n=146$).



Extended Data Fig. 3 | Fine-scale population structure. The point cloud at the top centre shows an alternative view of the UMAP result from Fig. 2b, with all ancient individuals coloured on the basis of analysis group. The framed panels surrounding the point cloud highlight particular ancestry clusters (as indicated), with labels and larger symbols corresponding to the median coordinates for the respective group. Similarly, the larger bottom panel shows median group coordinates for the large central point cloud, which includes the vast majority of European individuals from the Bronze Age onwards.

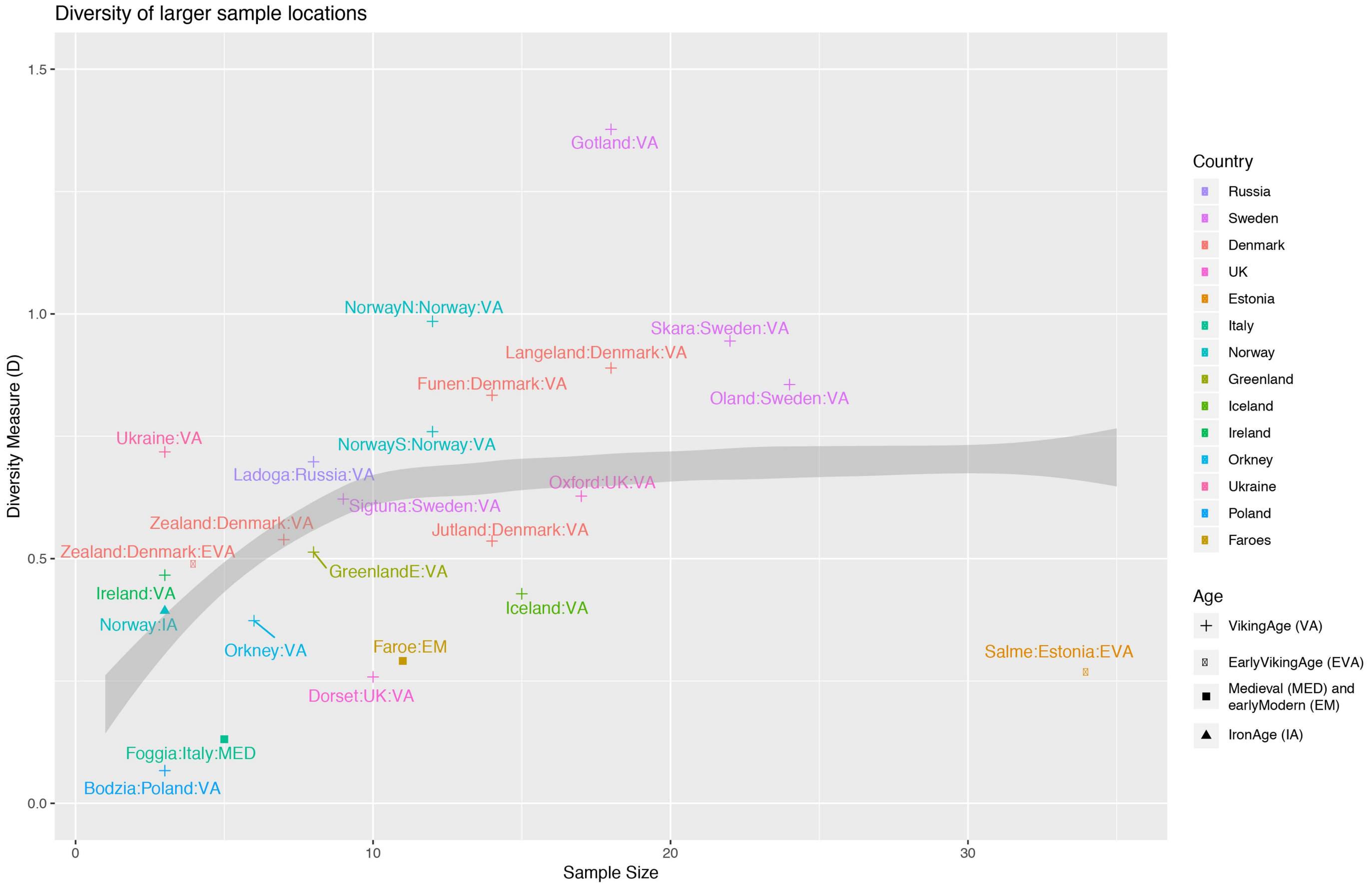


Extended Data Fig. 4 | Ancestry modelling for distal sources. a, Contrasting allele-sharing between Anatolian farmers (Barcin_EN) and Steppe pastoralists (Yamnaya_EBA) for European individuals from the Bronze Age and later. Violin plots showing distributions of statistics f_4(YRI,test individual;Barcin_EN,Yamnaya_EBA) for $n = 515$ individuals with a minimum of 1,000,000 SNPs with genotypes and groups with at least 2 such individuals. **b**, Ancestry proportions of analysis groups from the Bronze Age and later inferred using qpAdm. Target groups were modelled using three distal sources representing European hunter-gatherer (Loschbour_M), Anatolian farmer (Barcin_EN) and Steppe pastoralist (Yamnaya_EBA) ancestry. Sample sizes for target groups can be found in Supplementary Table 10. Error bars indicate standard error obtained from qpAdm. **c**, Ancestry proportions of analysis groups for which the three-source model was rejected using qpAdm ($P < 0.05$). Target groups were modelled including one additional distal source representing either Steppe hunter-gatherer (Botai_EBA), Caucasus hunter-gatherer (CaucasusHG_M) or East-Asian-related (XiongNu_IA) ancestry.

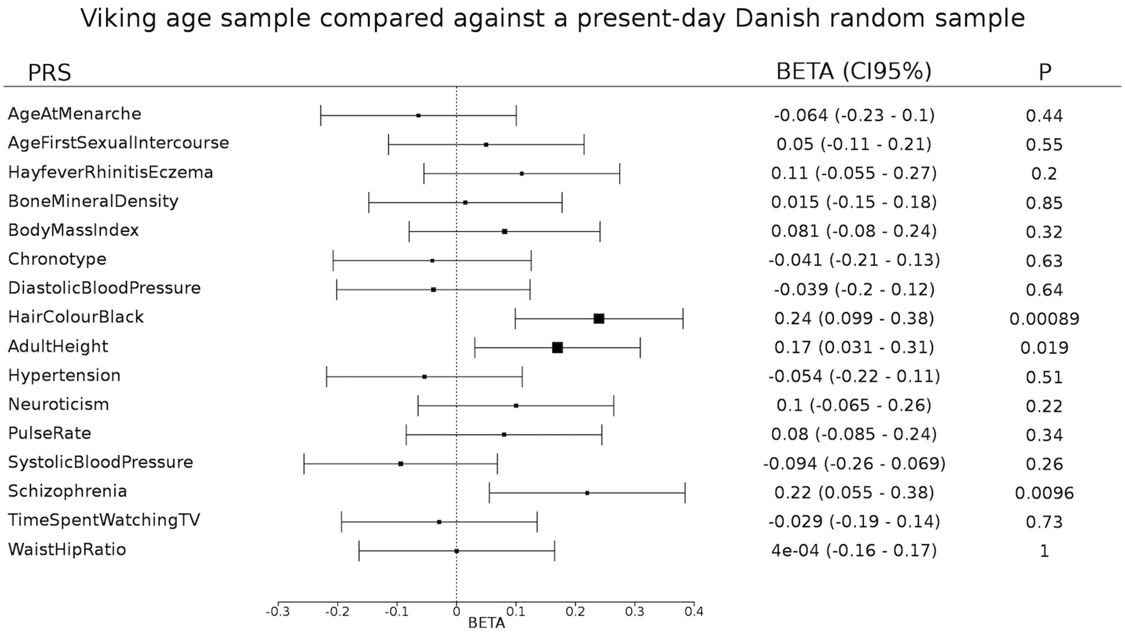


Extended Data Fig. 5 | See next page for caption.

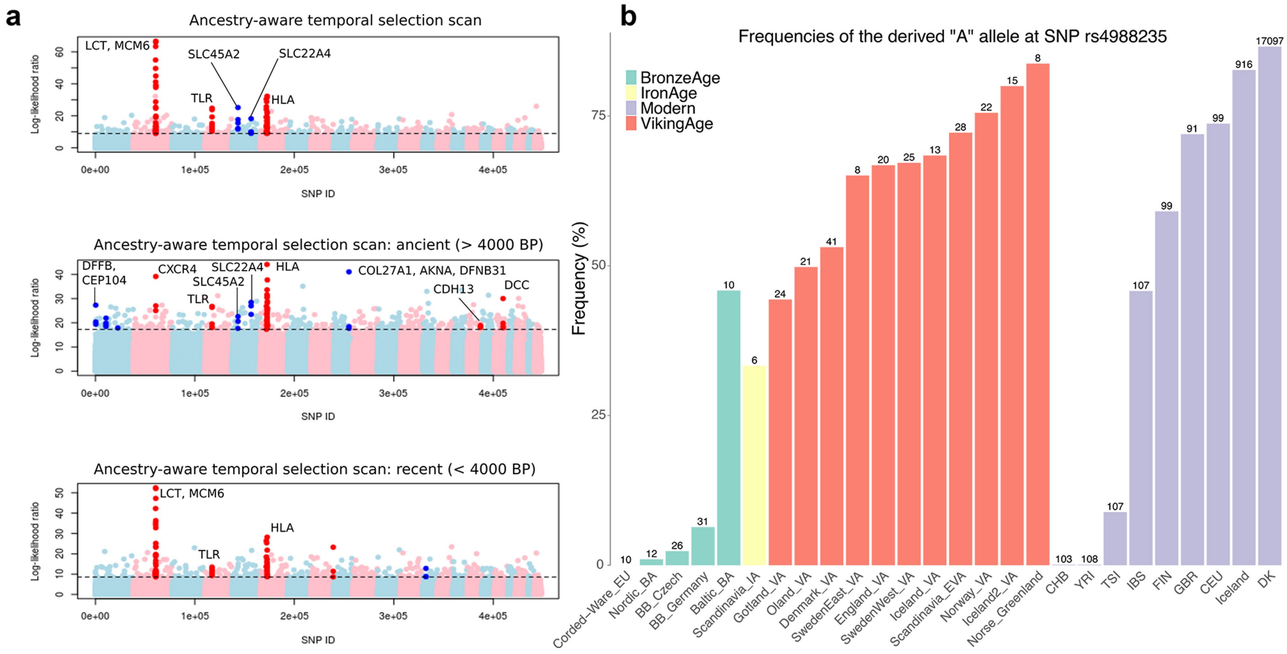
Extended Data Fig. 5 | Ancestry modelling for proximate sources. **a**, Testing for continuity between European Iron Age and later Viking Age and Medieval groups. Coloured squares depict whether a particular target group (row) can be modelled using a single source group (column). *P* values for f_4 rank of 0 (corresponding to a single source group) were obtained using qpAdm with a set of 15 outgroups, which included European Bronze Age groups that preceded the source groups. Sample sizes for target groups can be found in Supplementary Table 12. **b**, Two-way admixture ancestry proportions of target groups for which a single source was rejected ($P \leq 0.05$). Target groups were modelled using additional proximate Bronze and Iron Age sources. Sample sizes for target groups can be found in Supplementary Table 13. For both **a** and **b**, only ancient groups containing at least 3 individuals with a minimum of 1,000,000 SNPs with genotypes are plotted. **c**, Contrasting allele-sharing between populations of present-day Denmark and other populations. Violin plots showing distributions of statistics f_4(YRI,test individual;panel population,Denmark) for $n = 489$ individuals with a minimum of 50,000 SNPs with genotypes and groups with at least 2 such individuals. Median values for distributions are indicated with horizontal lines.



Extended Data Fig. 6 | Ancestry diversity of different population groups. Diversity of different labels (that is, sample locations combined with historical age) are shown as a function of their sample size. The diversity measure is the Kullback–Leibler divergence from the label means, capturing the diversity of a group with respect to the average of that group (see Supplementary Note 11 for details). Larger values are more diverse, although a dependence on sample size is expected. The simulation expectation for the best fit to the data (0 = 0.2) is shown.



Extended Data Fig. 7 | Polygenic risk scores. Polygenic risk scores (PRS) for 16 complex human traits in 148 Viking Age samples from Denmark, Sweden and Norway, compared against a reference sample of 20,551 Danish-ancestry individuals randomly drawn from all individuals born in Denmark in 1981–2005. The PRS is in each case based on allelic effects for >100 independent genome-wide significant SNPs from recent genome-wide association studies of the respective traits and standardised to a mean of 0 and standard deviation of 1 in the entire sample. Difference in PRS was estimated in a linear regression correcting for sex and 25 principal components of overall genetic structure. The plotted BETA indicates the coefficient for the test-group (Viking Age sample) PRS compared to that of the Danish comparison sample, with error bars indicating the 95% confidence interval of BETA, and *P* indicating the two-tailed *P* value of the corresponding *t*-test (not corrected for number of tests). Only PRS for black hair colour is significantly different between the groups after taking account of multiple testing.



Extended Data Fig. 8 | Positive selection in Europe. a, Manhattan plots of the likelihood ratio scores in favour of selection looking at the entire 10,000-year period (top, general scan), the period up to 4,000 years before present (middle, ancient scan) and the period from 4,000 years before present up to the present day (bottom, recent scan). The highlighted SNPs have a score larger than the 99.9% quantile of the empirical distribution of log-likelihood ratios, and have at least two neighbouring SNPs (± 500 kb) with a score larger than the same quantile. *n* = 1,185 genomes are used in the selection scan. **b**, Frequencies of the derived A allele rs4988235 SNP responsible for lactase persistence in humans for different Viking Age groups, present-day populations from the 1000 Genomes Project as well as relevant Bronze Age population panels. The numbers at the top of the bars denote the sample size on which the allele frequency estimates are based.

nature research

Corresponding author(s): Eske Willerslev

Last updated by author(s): May 3, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No specific software was used for data collection. All software used in this study is listed below.
Data analysis	All software used in this work is publicly available. Corresponding publications are cited in the main text and supplementary material. List of software and respective versions: CASAVA v1.8.2 AdapterRemoval v2.1.3 bwa v0.7.10 bwa mem 0.7.10 picard tools v1.127 bamUtil v1.0.14 samtools v1.3.1 GATK v3.3.0 pysam 0.7.4 (python module) bedtools 2.27.1 mapDamage2.0 contamMix v1.0-5 SHRiMP 2.2.3 YFitter v0.2 Haplogrep 2.0 FineSTRUCTURE ANGSD v0.915 IBDseq v.r1206 PRANK v.150803 pathPhynder BEASTv1.8.2

schmutzi v.1.5.4
Yleaf v2
admixtools v4.1
NGSrelate v.1
GLOBETROTTER
READ
PRIMUS v1.9
plink and v1.9
ADMIXTURE v1.3
RAxML-8.1.15
SnpEff
SPAdes-3.9.0
R 3.2.3
Ohana
Beagle v4.1
python 2.7.12
perl v5.22.1
CALIB
FigTree v.1.4.4

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequence data are available at the European Nucleotide Archive under accession number PRJEB37976.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not rely on statistical methods to predetermine sample sizes. Sample sizes in ancient population genetic studies are limited by the number of samples yielding endogenous DNA proportions amenable to whole genome sequencing.
Data exclusions	We selected 442 samples for whole-genome sequencing, out of all (n=528) screened samples, based on their endogenous content and low contamination estimates. These criteria are described in detail in Supplementary Notes 2 and 3. Furthermore, closely related individuals were excluded from analyses requiring population allele frequencies.
Replication	We did not attempt to specifically replicate experimental findings. But we note that samples from the same population carry similar genetic signatures. Moreover, genome-wide data allows for the analysis of multiple realisations of the sample history, by studying hundreds of thousands of SNP sites.
Randomization	We did not implement any randomization as no experimental groups or effect sizes were measured in this study.
Blinding	No blinding techniques were implemented, as experimental group assignment is not relevant for population genetic studies of this kind.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Suppression of proteolipid protein rescues Pelizaeus–Merzbacher disease

https://doi.org/10.1038/s41586-020-2494-3

Received: 29 December 2018

Accepted: 24 June 2020

Matthew S. Elitt[1], Lilianne Barbar[1], H. Elizabeth Shick[1], Berit E. Powers[2], Yuka Maeno-Hikichi[1], Mayur Madhavan[1], Kevin C. Allan[1], Baraa S. Nawash[1], Artur S. Gevorgyan[1], Stevephen Hung[1], Zachary S. Nevin[1], Hannah E. Olsen[1], Midori Hitomi[3], Daniela M. Schlatzer[4], Hien T. Zhao[2], Adam Swayze[2], David F. LePage[1], Weihong Jiang[1], Ronald A. Conlon[1], Frank Rigo[2] & Paul J. Tesar[1]✉

Mutations in *PLP1*, the gene that encodes proteolipid protein (PLP), result in failure of myelination and neurological dysfunction in the X-chromosome-linked leukodystrophy Pelizaeus–Merzbacher disease (PMD)[1,2]. Most *PLP1* mutations, including point mutations and supernumerary copy variants, lead to severe and fatal disease. Patients who lack *PLP1* expression, and *Plp1*-null mice, can display comparatively mild phenotypes, suggesting that *PLP1* suppression might provide a general therapeutic strategy for PMD[1,3–5]. Here we show, using CRISPR–Cas9 to suppress *Plp1* expression in the *jimpy* (*Plp1*^{jp}) point-mutation mouse model of severe PMD, increased myelination and restored nerve conduction velocity, motor function and lifespan of the mice to wild-type levels. To evaluate the translational potential of this strategy, we identified antisense oligonucleotides that stably decrease the levels of *Plp1* mRNA and PLP protein throughout the neuraxis in vivo. Administration of a single dose of *Plp1*-targeting antisense oligonucleotides in postnatal *jimpy* mice fully restored oligodendrocyte numbers, increased myelination, improved motor performance, normalized respiratory function and extended lifespan up to an eight-month end point. These results suggest that *PLP1* suppression could be developed as a treatment for PMD in humans. More broadly, we demonstrate that oligonucleotide-based therapeutic agents can be delivered to oligodendrocytes in vivo to modulate neurological function and lifespan, establishing a new pharmaceutical modality for myelin disorders.

PMD (Online Mendelian Inheritance in Man (OMIM) ID: 312080) is a fatal, X-linked leukodystrophy characterized by extensive loss of myelinating oligodendrocytes in the central nervous system (CNS). Mutations in the *PLP1* gene, which encodes the highly conserved four-transmembrane-domain oligodendrocyte protein PLP, cause PMD[1,2]. Symptoms typically present at birth or in childhood, and include a constellation of nystagmus, spasticity, hypotonia and cognitive dysfunction, leading to early death, often before adulthood. Preclinical efforts to extend lifespan have had only limited success, and no therapy has shown efficacy in patients[6–13].

Most patients with PMD have *PLP1*-duplication mutations, which cause overexpression of otherwise normal PLP protein[1,2]. However, hundreds of distinct PMD-causative point mutations, which result in abnormal PLP protein, have also been identified. Notably, rare patients who lack *PLP1* expression display symptoms that are delayed and milder compared with those with more severe duplications or point mutations[3–5]. These *PLP1*-null patients can live for 40–60 years, do not develop spastic paraparesis until the second or third decade of life, and maintain intact cognition until the third or fourth decade of life (Supplementary Table 1), possibly owing to a lack of cellular stress responses and oligodendrocyte death triggered by excess or abnormal PLP[1,2,12,14,15].

This clinical landscape suggests several opportunities for therapeutic development. Specifically, reducing *PLP1* expression to normal levels in patients with gene duplications would be expected to be curative. More broadly, the milder presentation of patients lacking *PLP1* implies a wide therapeutic window for titrating *PLP1* expression, which could be leveraged to restore functional oligodendrocytes in patients with point mutations that generate abnormal PLP. Here we demonstrate therapeutic *Plp1* suppression using germline- and postnatal-based approaches in a mouse model of PMD that expresses abnormal PLP.

Germline suppression of *Plp1* in PMD mice

To test whether *Plp1* suppression provides a generalizable therapeutic approach for PMD, we used the *jimpy* (*Plp1*^{jp}) mouse model of PMD, which expresses abnormal PLP and recapitulates the cellular, molecular and neurologic features seen in severe PMD. We targeted *Plp1* with

---

[1]Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA. [2]Ionis Pharmaceuticals, Carlsbad, CA, USA. [3]Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA. [4]Case Center for Proteomics and Bioinformatics, Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, OH, USA. ✉e-mail: paul.tesar@case.edu

CRISPR[16,17], using single guide RNAs (sgRNAs) with high on-target, germline cutting efficiency (Supplementary Table 2), to generate a CRISPR-modified *jimpy* (CR-*impy*) founder with a complex deletion in *Plp1* (Fig. 1a, Extended Data Fig. 1a–c). Before subsequent analyses, rare off-target mutations were eliminated by back-crossing (Extended Data Fig. 1b, d, e).

Whereas *jimpy* mice showed severe tremor, ataxia, seizures (lasting more than 30 s) and death by the third postnatal week, CR-*impy* mice exhibited a 21-fold increase in lifespan (mean survival 489 and 23 days for CR-*impy* and *jimpy* mice, respectively) with no evidence of tremor, ataxia or seizures up to the terminal end point of 18 months of age (Fig. 1b, Supplementary Data 1, Supplementary Videos 1, 2). The level of *Plp1* transcript in CR-*impy* mice was reduced by 61–74% relative to wild type in multiple CNS regions at 6 months of age (Extended Data Fig. 2a), with undetectable levels of PLP protein (Supplementary Table 3).

To explore the effects of germline *Plp1* suppression on cellular pathology, we assessed markers of oligodendrocyte lineage and neuroinflammation. The mature myelin marker myelin basic protein (MBP) was grossly and stably restored to near wild-type levels throughout the neuraxis in CR-*impy* mice (Fig. 1c). In contrast to the almost complete absence of *Mbp* expression in *jimpy* mice, CR-*impy* mice demonstrated substantially increased transcript (83–91% of wild type at 6 months of age) and protein (40–95% and 114–130% of wild type at 3 weeks and 6 months of age, respectively) in multiple CNS regions (Extended Data Fig. 2b–d, Supplementary Data 2a, b, 3–5). Quantification of myelin regulatory factor (MyRF)-positive oligodendrocytes showed their complete restoration throughout multiple CNS regions in CR-*impy* mice (94–117% and 89–126% of wild type at 3 weeks and 6 months of age, respectively), in contrast to their depletion in *jimpy* mice (36–59% of wild type at 3 weeks of age) (Fig. 1d, e, Supplementary Data 3–5). The glial lineage marker SOX10, which is expressed by oligodendrocytes and oligodendrocyte progenitor cells (OPCs), showed no differences across these genotypes (Fig. 1d, e, Supplementary Data 3–5). CR-*impy* mice showed minimal evidence of astrogliosis or microglial activation up to 6 months of age, in contrast to elevated neuroinflammatory markers in *jimpy* mice[18] (Extended Data Fig. 3a–d, Supplementary Data 3–5).

To investigate *Plp1* suppression in oligodendrocytes isolated from cell-extrinsic developmental or inflammatory cues, we generated and characterized induced pluripotent stem cell lines (Extended Data Fig. 4a, b), which were differentiated to oligodendrocytes in vitro. Notably, CR-*impy* lines showed cell-type-specific rescue in oligodendrocyte number and arborized morphology relative to *jimpy* lines (Extended Data Fig. 4c–g). Collectively, these data confirm that *Plp1* suppression has a cell-intrinsic effect on oligodendrocytes that is sufficient to rescue *jimpy* cellular phenotypes.

To assess the effect of germline *Plp1* suppression on myelination, we quantified electron micrograph data. In contrast to nearly absent myelination in *jimpy* mice, CR-*impy* mice showed a marked increase in myelinated axons throughout the neuraxis, reaching nearly 50% of that in wild-type mice by 3 weeks of age, with stability up to 18 months of age (Fig. 1f–h). Myelin sheaths in CR-*impy* mice showed incomplete compaction compared with those in the wild-type mice, consistent with the role of PLP in myelin ultrastructure[5,19] (Fig. 1f, g). To determine whether myelin in CR-*impy* mice was functional, we measured compound action potential speed in the optic nerve. At 3 weeks of age, we found a significant increase in conduction velocity in CR-*impy* mice relative to *jimpy* mice (Fig. 1i) (reaching approximately 55% of that in the wild type), which was well-correlated with the level of myelination in CR-*impy* mice (approximately 35% of the wild-type level) (Fig. 1j). Notably, CR-*impy* and wild-type mice showed similar conduction velocities at 6 months of age (Fig. 1i).

To determine whether restored myelin altered complex motor function, we used longitudinal open-field and rotarod testing. Overall locomotion was decreased in *jimpy* mice, but similar between CR-*impy* and wild-type mice across all time points (Fig. 1k). Rotarod testing revealed that CR-*impy* mice showed similar performance to wild-type mice up to 6 months of age, whereas *jimpy* mice exhibited significant impairment. At 18 months of age, the CR-*impy* mice displayed slightly reduced performance (Fig. 1l), potentially reflecting late-onset neuronal phenotypes[4]. Together, these results establish that germline suppression of *Plp1* restores oligodendrocytes, functional myelin and lifespan in *jimpy* mice.

In vivo suppression of oligodendrocyte transcripts

After validating *Plp1* as a therapeutic target for PMD using germline suppression, we pursued a clinically translatable strategy for in vivo, postnatal *Plp1* suppression using newer-generation antisense oligonucleotides (ASOs). These ASOs, distinguished by their highly efficient modulation of target transcripts in the CNS with multi-month in vivo half-lives, underlie several therapies for fatal neuronal-based disorders[20–22]; however their ability to target the oligodendrocyte lineage in vivo was unknown. To establish their therapeutic potential for this lineage, we administered well-characterized ASOs targeting *Hdac2* to adult wild-type mice by intracerebroventricular (ICV) injection. HDAC2 protein is localized to the nucleus and enables clear visualization of target suppression; ICV injection of ASOs targeting *Hdac2* resulted in a substantial reduction of HDAC2 levels in OPCs and oligodendrocytes (Fig. 2a, b). Next we identified two independent ASOs targeting the fifth intron (ASO*Plp1*.a) and 3′-untranslated region (UTR) (ASO*Plp1*.b) of *Plp1* (Fig. 2c), along with a non-targeting ASO control (ASOctr), which we administered to wild-type mice. These ASOs showed dose-dependent suppression of *Plp1* transcript (up to 90% and 98% suppression in neonatal and adult wild-type mice, respectively) and PLP protein (up to 63% in neonatal wild-type mice) in multiple CNS regions (Fig. 2d–g, Extended Data Fig. 5a). They also showed widespread distribution across the neuraxis, did not exhibit off-target effects on non-*Plp1* transcripts, did not activate glial cells and did not alter levels of MBP protein in wild-type mice (Extended Data Figs. 5b–h, 6a, b, 7a, Supplementary Data 7a).

Postnatal *Plp1* suppression in PMD mice

We evaluated the therapeutic effect of *Plp1*-targeting ASOs on the severe *jimpy* phenotype using a single ICV injection at birth (Fig. 3a). *Jimpy* mice injected with ASO*Plp1*.a or ASO*Plp1*.b exhibited increases in lifespan of approximately 12-fold and 11-fold, respectively, compared with *jimpy* mice injected with ASOctr (mean survival of 20 (ASOctr), 239 (ASO*Plp1*.a) and 217 days (ASO*Plp1*.b)) up to a predetermined terminal end point of 8 months of age (Fig. 3b, Supplementary Data 6, Supplementary Videos 3, 4).

MBP expression was grossly increased in *jimpy* mice treated with ASO*Plp1*.a or ASO*Plp1*.b relative to those treated with ASOctr up to eight months of age, without additional ASO dosing (Fig. 3c, d). Levels of *Mbp* transcript and MBP protein were significantly increased across the neuraxis in *jimpy* mice treated with ASO*Plp1*.a or ASO*Plp1*.b compared with those treated with ASOctr (up to a 39-fold increase in MBP protein), along with a concomitant reduction in apoptotic cells (Extended Data Fig. 7b–e, Supplementary Data 7b, 8–10). MyRF-positive oligodendrocytes were substantially depleted in *jimpy* mice treated with ASOctr but were restored throughout the neuraxis at 3 weeks of age in *jimpy* mice treated with ASO*Plp1*.a or ASO*Plp1*.b (81–101% of the level in wild-type mice treated with ASOctr) (Fig. 3e, Supplementary Data 8–10). These trends were further validated by examining OLIG2 and CC1 double-positive oligodendrocytes (Extended Data Fig. 7f, Supplementary Data 11–13). There were similar levels of SOX10- and OLIG2-positive oligodendrocyte lineage cells across groups (Fig. 3f, Extended Data Fig. 7g, Supplementary Data 8–13). Levels of PDGFRα and OLIG2 double-positive OPCs were similar in wild-type mice treated with ASOctr and in *jimpy* mice treated with ASO*Plp1*.a and ASO*Plp1*.b,

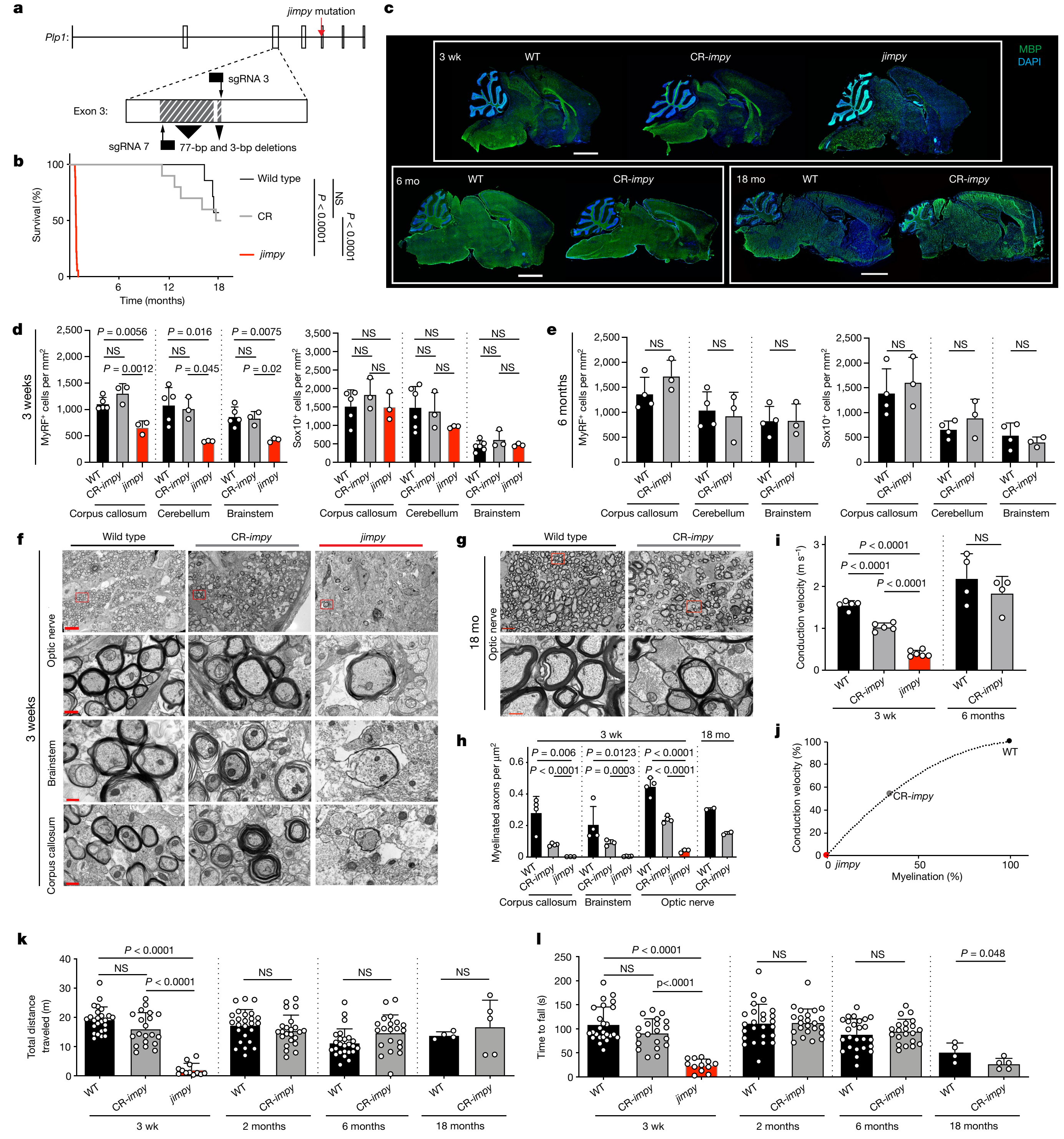


Fig. 1 | Germline *Plp1* suppression in *jimpy* rescues lifespan and restores functional myelin. **a**, Schematic of CRISPR *Plp1* targeting in *jimpy*. Black arrowheads indicate predicted sgRNA cutting sites. Hashed boxes show the CR-*impy* 80-base pair (bp) complex deletion (Extended Data Fig. 1a). **b**, Kaplan–Meier plot comparing lifespans between genotypes. $n = 25$ (wild type (WT)), 23 (CR-*impy*) and 18 (*jimpy*) mice. *P* values calculated using log-rank test. **c**, Immunohistochemistry of whole-brain sagittal sections of 3-week-old (wk), 6-month-old (mo) and 18 month-old wild-type (WT), CR-*impy* and *jimpy* mice showing MBP (green) and nuclei (DAPI, blue). Scale bars, 2 mm. **d**, **e**, Quantification of MyRF$^+$ and SOX10$^+$ cells at three weeks (**d**) and six months (**e**) of age. $n = 3$–6 mice. Representative source images are presented in Supplementary Data 3–5. **f**, **g**, Electron micrographs showing myelination at 3 weeks (**f**) and 18 months (**g**) of age. Higher magnification of the red boxed area shown in the next row. Scale bars, 5 µm (top row) and 0.5 µm (other rows). **h**, Quantification of myelinated axons at 3 weeks ($n = 3$–4 mice) and 18 months ($n = 2$ mice) of age. *P*-values calculated with unpaired, two-sided *t*-tests. **i**, Optic nerve conduction velocities at 3 weeks ($n = 5$–6 mice) and 6 months ($n = 4$ mice). **j**, Polynomial trend line illustrating conduction velocity versus brain myelination in CR-*impy* relative to minimum–maximum scaling of values from *jimpy* and wild-type mice. Data from three-week time point of **h** and **i**, with same *n*. **k**, **l**, Accelerating rotarod (**k**) or open-field (**l**) performance. $n = 25$ (WT), 20 (CR-*impy*) and 12 (*jimpy*) mice at 3 weeks of age; $n = 25$ (WT), 23 (CR-*impy*) mice at 2 months of age; $n = 25$ (WT), 21 (CR-*impy*) mice at 6 months of age; and $n = 4$ (WT), 5 (CR-*impy*) mice at 18 months of age. Biological replicates (individual mice) indicated by open circles. Data are mean ± s.d. *P* values calculated using one-way analysis of variance (ANOVA) with Tukey's correction at three weeks or two-way, unpaired two-sided *t*-test at later time points, except where indicated. *P* values shown for $P < 0.1$, otherwise not significant (NS).

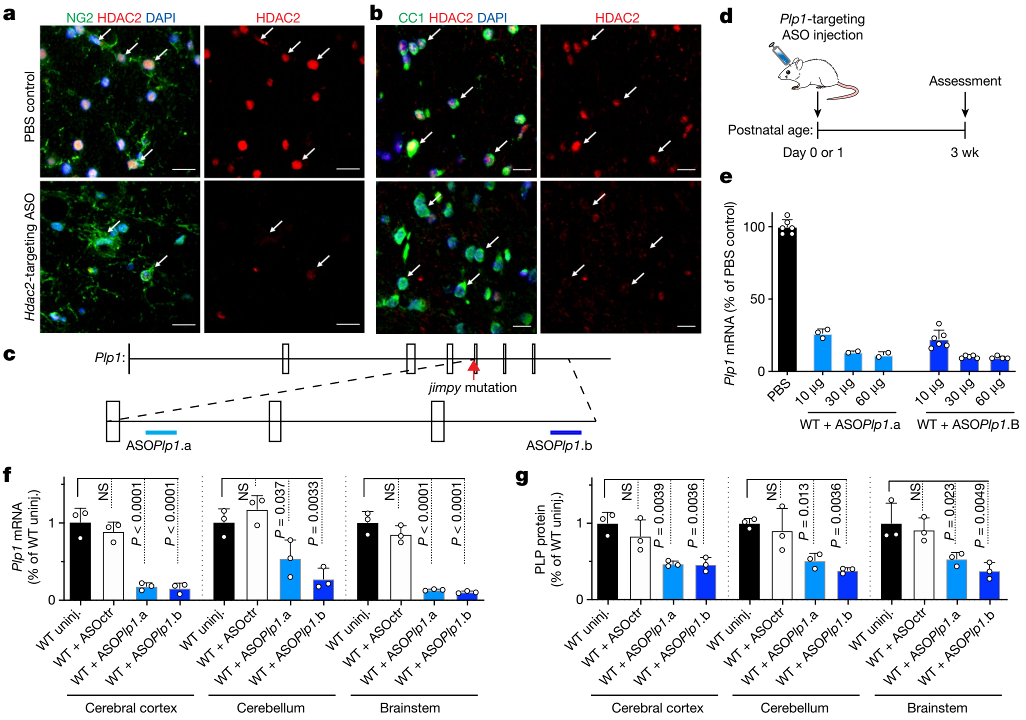


Fig. 2 | Efficient ASO-mediated transcript suppression in OPCs and oligodendrocytes in vivo. a, **b**, Immunostaining of HDAC2$^+$ (red) and NG2$^+$ OPCs (green; arrows) in the spinal cord (**a**) or CC1$^+$ oligodendrocytes (green; arrows) in the corpus callosum (**b**) from eight-week-old wild-type mice injected with PBS control or *Hdac2*-targeting ASO, two weeks after injection. Scale bars, 20 µm. **c**, Depiction of *Plp1* pre-mRNA, showing the approximate binding locations of ASO*Plp1*.a and ASO*Plp1*.b in intron 5 and the 3′ UTR, respectively. **d**, Schematic of the design of ASO experiments in this figure. **e**, Quantitative PCR with reverse transcription (RT–qPCR) data showing wild-type *Plp1* transcript levels in the spinal cord, 3 weeks after injection with the indicated ASO doses (10 µg, 30 µg or 60 µg) or PBS controls at postnatal day 1 ($n = 2$–6 mice). **f**, **g**, RT–qPCR data showing the levels of *Plp1* transcript (**f**) and western blot data showing the levels of PLP protein (**g**), 3 weeks after ASO injection (30 µg dose) at birth in wild-type mice ($n = 3$ mice). Uninj., uninjected. Individual data points represent the mean value of four technical replicates for each biological replicate (**e**, **f**) or independent biological replicates (**g**). Biological replicates (individual mice) indicated by open circles. Data are mean ± s.d. *P* values calculated using one-way ANOVA with Dunnett's correction. *P* values shown for $P < 0.1$, otherwise not significant. See Supplementary Data 4 for full western blot source images.

but were increased in *jimpy* mice treated with ASOctr (Extended Data Fig. 7h, Supplementary Data 11–13), suggesting a *jimpy*-specific compensatory increase in progenitors[23]. Myelinated axons were significantly increased throughout the neuraxis in *jimpy* mice treated with ASO*Plp1*.a or ASO*Plp1*.b relative to those treated with ASOctr at 3 weeks of age (approximately 5–6-fold and 12–15-fold higher in the corpus callosum and brainstem, respectively) (Fig. 3g, h). Although oligodendrocyte numbers were fully restored, myelination in these mice was only about 10% of the level in wild-type mice treated with ASOctr at 3 weeks of age and persisted up to the 8-month end point, albeit with less compaction (Fig. 3g, h, Extended Data Fig. 8a, b).

Notably, *jimpy* mice treated with ASO*Plp1*.a or ASO*Plp1*.b showed only mild *jimpy* phenotypes, including markedly reduced tremor and occasional short-duration seizures (less than 15 s), and appeared outwardly normal otherwise, including in the ability to breed (Supplementary Data 6). Although rotarod performance of *jimpy* mice was only variably and partially improved with ASO*Plp1*.a or ASO*Plp1*.b treatment (to a maximum of 36% of wild-type performance), overall locomotion was restored to wild-type levels across multiple time points (Fig. 4a, b). To assess whether myelin might contribute to these functional improvements, we measured compound action potential speed in the optic nerve. At three weeks of age, we found a modest but significant increase in conduction velocity in *jimpy* mice treated with ASO*Plp1*.b versus ASOctr (Fig. 4c), representing about 17% of the level in wild-type mice treated with ASOctr; this result corresponds directly with the level of myelination relative to the wild-type control mice (about 10%) (Fig. 4d). Together, these data demonstrate that a single postnatal administration of *Plp1*-targeting ASOs in *jimpy* mice elicits a sustained reduction in *Plp1* expression that restores oligodendrocytes and increases functional myelin, with improvements in motor performance and lifespan.

Respiratory distress and dysfunction has been associated with premature death in animal models of PMD and in patients with the disease[24–27], which is notable given the marked increase in survival of *jimpy* mice treated with ASO*Plp1*.a or ASO*Plp1*.b in light of the relatively modest increases in myelin globally, with the highest levels consistently observed in the brainstem (Fig. 3g, h, Extended Data 7b, c). Notably, brainstem respiratory control centres alter breathing patterns in response to physiological derangements seen during hypoxia or hypercapnia. Seizures, as observed in *jimpy* mice from around the third postnatal week, can trigger such derangements (Fig. 4e) and, when coupled with a reduced capacity to achieve homeostasis, could be lethal.

To investigate whether respiratory function is a therapeutic component of *Plp1*-targeting ASOs, we used plethysmography to measure minute ventilation in normal air, hypercapnic (5% CO_2) and hypoxic (10.5% O_2) conditions (Supplementary Data 14). When transitioned from normal air to either hypercapnic or hypoxic environments, *jimpy* mice treated with ASOctr exhibited high variability in minute ventilation, indicative of dysfunctional respiratory control (Fig. 4f), whereas those treated with ASO*Plp1*.b showed less variability and responses more similar to those of wild-type control mice (Fig. 4f–j). Specifically, *jimpy* mice treated with ASOctr showed weak compensatory decreases in minute ventilation when exposed to hypercapnic conditions relative to wild-type mice treated with ASOctr, which were restored in *jimpy* mice treated with ASO*Plp1*.b (Fig. 4g, h). During early transition to hypoxia, wild-type mice treated with ASOctr and *jimpy* mice treated

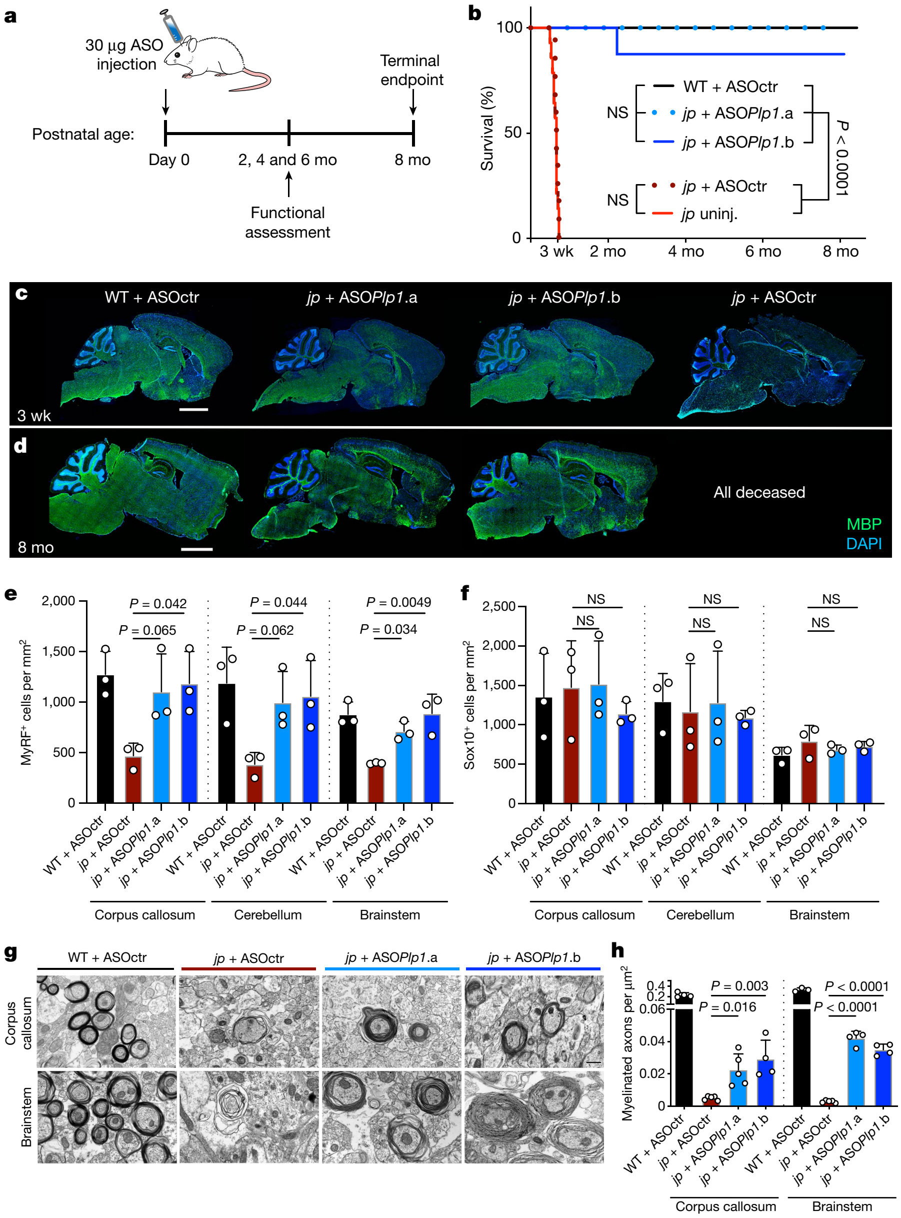


Fig. 3 | Postnatal delivery of *Plp1*-targeted ASOs rescues lifespan and oligodendrocytes with partial restoration of myelin in *jimpy* mice. **a**, Schematic of ASO experimental design used in Figs. 3, 4. **b**, Kaplan–Meier plot showing the lifespan of contemporaneous wild-type mice treated with ASOctr ($n = 12$) and *jimpy* (*jp*) mice uninjected ($n = 14$) or injected with ASOctr, ASO*Plp1*.a ($n = 5$) or ASO*Plp1*.b ($n = 5$). *P* values calculated using the log-rank test. See Supplementary Data 6 for source metadata. **c**, **d**, Immunohistochemical images of 3-week-old (**c**) and 8-month-old (**d**) whole-brain sagittal sections showing MBP (green) and DAPI (blue) staining. Scale bars, 2 mm. See Supplementary Data 8–10 for higher magnification. **e**, **f**, Quantification of MyRF$^+$ oligodendrocytes (**e**) and SOX10$^+$ glial lineage cells (**f**) at 3 weeks of age ($n = 3$ mice). For representative source images, see Supplementary Data 8–10. **g**, **h**, Electron micrographs (**g**) and quantification (**h**) of myelinated axons at 3 weeks of age ($n = 3$–5 mice). Scale bar, 0.5 µm. Biological replicates (individual mice) indicated by open circles. Data are mean ± s.d. *P* values calculated using one-way ANOVA with Dunnett's correction, except where indicated. *P* values shown for $P < 0.1$, otherwise not significant.

with ASO*Plp1*.b demonstrated similar compensatory increases in minute ventilation, whereas *jimpy* mice treated with ASOctr showed a blunted response (Fig. 4g, i). In extended hypoxia, *jimpy* mice treated with ASOctr showed an exaggerated decrease in minute ventilation relative to wild-type controls, which was restored when they were treated with ASO*Plp1*.b (Fig. 4j). Of note, during this hypoxic challenge, 38% of *jimpy* mice treated with ASOctr died spontaneously, whereas 100% of those treated with ASO*Plp1*.b and wild-type mice treated with ASOctr survived (Fig. 4k). Together, these results suggest that dysregulated control of respiration is a component of the *jimpy* phenotype and potentially underlies the premature mortality that occurs coincident with the onset of seizures, and can be partially rescued by suppression of *Plp1*.

Discussion

In summary, we have validated a clinically feasible therapeutic strategy for PMD using a mutation-agnostic approach based on suppression of PLP. We demonstrate that suppression of *Plp1* expression using CRISPR–Cas9 in the germline or postnatal ASO results in rescue of major PMD phenotypes in a mouse model of severe PMD. Furthermore, we establish that oligonucleotide-based drugs, delivered postnatally, can modulate a disease target in oligodendrocytes and restore both functional myelin and lifespan in mice with a fatal genetic disorder.

This study provides foundational data for the development of clinically relevant ASO technology to achieve postnatal reduction of *PLP1*.

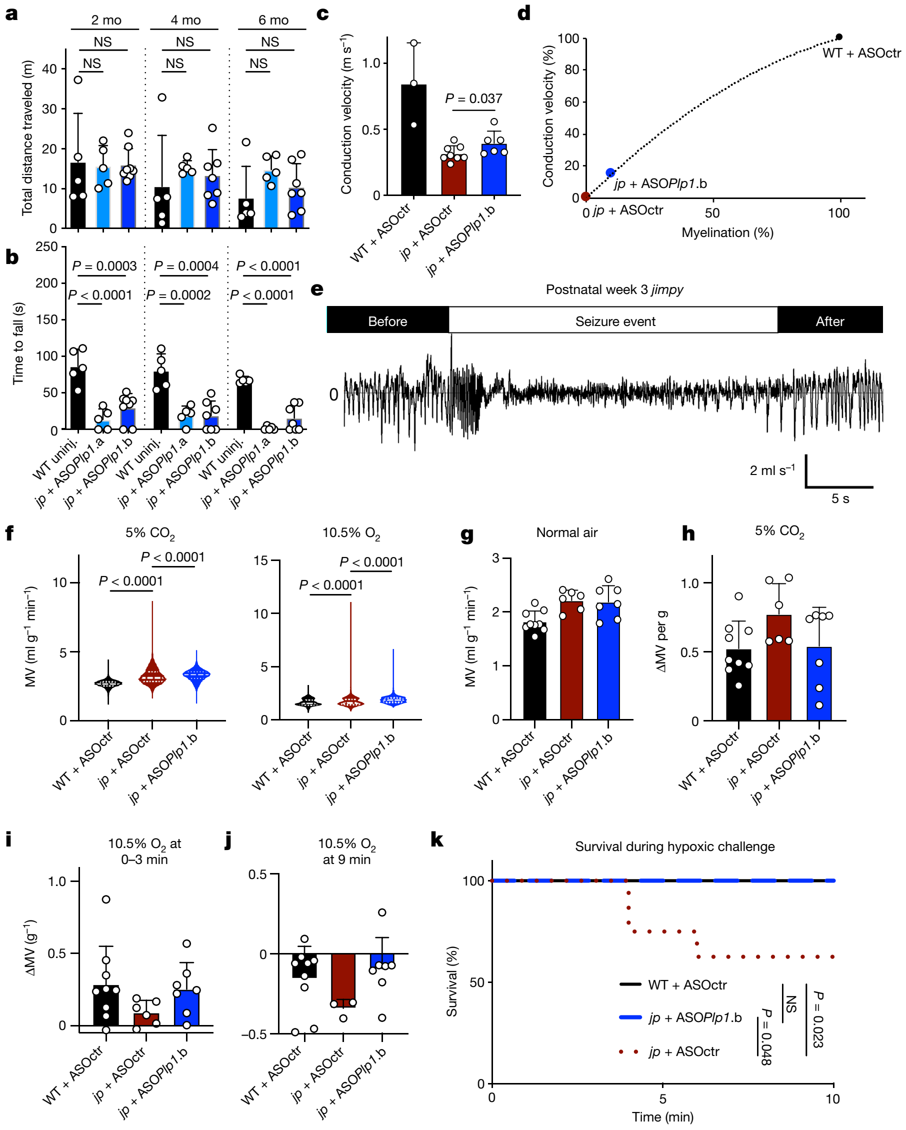


Fig. 4 | ASO-mediated *Plp1* suppression in *jimpy* leads to functional myelin, improved control of respiratory function and prevention of hypoxia-induced mortality. a, b, Performance in open-field testing (**a**) and accelerating rotarod (**b**) ($n = 5$–8 mice). Raw data is presented in Supplementary Data 6. *P* values calculated using one-way ANOVA with Dunnett's correction. **c**, Optic nerve conduction velocity at 3 weeks of age ($n = 3$ (WT + ASOctr), 8 (*jimpy* + ASOctr) and 4 (*jimpy* + ASO*Plp1*.b) mice). *P* values calculated using one-sided, unpaired *t*-test. **d**, Polynomial trend line illustrating conduction velocity versus brain myelination in *jimpy* + ASO*Plp1*.b to minimum–maximum scaling of values from *jimpy* and wild-type mice treated with ASOctr. Source data from **c** and Fig. 3h with same number of samples. **e**, Trace of a *jimpy* seizure during hypercapnic challenge (*y*-axis, respiratory flow rate). **f**, Minute ventilation (MV) in ml g^{-1} min^{-1} and per gram body weight (ΔMV per g) in 5% CO_2 (hypercapnia) and 10.5% O_2 (hypoxia), representing all repeated measurements from $n = 9$ (WT + ASOctr), 6 (*jimpy* + ASOctr) and 7 (*jimpy* + ASO*Plp1*.b) mice. Violin plots indicate median (centre white lines) and quartiles (border white lines). *P* values calculated using Brown and Forsythe's test. **g**–**j**, Minute ventilation per body weight (MV g^{-1}) in normal air (**g**), 15–30 min after transitioning from normal air to 5% CO_2 (**h**), 0–3 min (**i**) and 8–9 min (**j**) after transitioning from normal air to 10.5% O_2. Post-mortality hypoxia data were not included. In **g**–**i**, $n = 9$ (WT + ASOctr), 6 (*jimpy* + ASOctr) and 7 (*jimpy* + ASO*Plp1*.b) mice; in **j**, $n = 9$ (WT + ASOctr), 3 (*jimpy* + ASOctr) and 7 (*jimpy* + ASO*Plp1*.b) mice. **k**, Kaplan–Meier plot showing survival during hypoxia. $n = 12$ (WT + ASOctr), 8 (*jimpy* + ASOctr) and 9 (*jimpy* + ASO*Plp1*.b) mice. *P* values calculated using log-rank test. Biological replicates (individual mice) indicated by open circles. Data are mean ± s.d., except where indicated. *P* values shown for $P < 0.1$, otherwise not significant.

Further preclinical development is needed to optimize dosage and timing, including treatment later in disease progression; nevertheless, our results highlight that even a single ASO injection can elicit a sustained phenotypic improvement relative to the natural history of the disease, even with restoration of about 10% of myelin relative to wild type. These data could reflect a previously unappreciated functional tolerance to incomplete myelination or may be indicative of a neuronal-supportive function of oligodendrocytes[28–30], levels of which were completely restored in *jimpy* mice injected with *Plp1*-targeting ASOs.

Complete elimination of mutant PLP could convert patients with severe PMD to a *PLP1*-null phenotype, characterized by milder disease that presents later, progresses slower and shows improved clinical outcomes[1,3–5]. Titration of abnormal or excessive PLP to a level that relieves cellular stress-mediated oligodendrocyte death but maintains the neuronal-supportive function of PLP[3–5,19] could potentially provide greater benefit. This strategy would be especially amendable to the 70% of PMD patients who have gene duplications leading to excess levels of normal PLP protein[1], as a reduction to wild-type levels of PLP expression may be curative.

Collectively, our studies, combined with the feasibility of ASO delivery to the human CNS and current safety data in other CNS indications, support advancement of *PLP1* suppression into the clinic as a therapeutic strategy with potential applicability across the spectrum of patients with PMD. More broadly, our data provide a framework for transcript modulation in oligodendrocytes to restore myelination in genetic and sporadic disorders of myelination.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2494-3.

1. Inoue, K. Pelizaeus–Merzbacher Disease: molecular and cellular pathologies and associated phenotypes. *Adv. Exp. Med. Biol*. **1190**, 201–216 (2019).
2. Wolf, N. I., van Spaendonk, R. M. L., Hobson, G. M. & Kamholz, J. in *Gene Reviews* (eds Adam, M. P. et al.) https://www.ncbi.nlm.nih.gov/books/NBK1182/ (1993).
3. Garbern, J. Y. et al. Patients lacking the major CNS myelin protein, proteolipid protein 1, develop length-dependent axonal degeneration in the absence of demyelination and inflammation. *Brain* **125**, 551–561 (2002).
4. Griffiths, I. et al. Axonal swellings and degeneration in mice lacking the major proteolipid of myelin. *Science* **280**, 1610–1613 (1998).
5. Klugmann, M. et al. Assembly of CNS myelin in the absence of proteolipid protein. *Neuron* **18**, 59–70 (1997).
6. Goldman, S. A., Nedergaard, M. & Windrem, M. S. Glial progenitor cell-based treatment and modeling of neurological disease. *Science* **338**, 491–495 (2012).
7. Gupta, N. et al. Neural stem cell engraftment and myelination in the human brain. *Sci. Transl. Med*. **4**, 155ra137 (2012).
8. Saher, G. et al. Therapy of Pelizaeus–Merzbacher disease in mice by feeding a cholesterol-enriched diet. *Nat. Med*. **18**, 1130–1135 (2012).
9. Wishnew, J. et al. Umbilical cord blood transplantation to treat Pelizaeus–Merzbacher disease in 2 young boys. *Pediatrics* **134**, e1451–e1457 (2014).
10. Tantzer, S., Sperle, K., Kenaley, K., Taube, J. & Hobson, G. M. Morpholino antisense oligomers as a potential therapeutic option for the correction of alternative splicing in PMD, SPG2, and HEMS. *Mol. Ther. Nucleic Acids* **12**, 420–432 (2018).
11. Li, H. et al. Gene suppressing therapy for Pelizaeus–Merzbacher disease using artificial microRNA. *JCI Insight* **4**, 125052 (2019).
12. Nobuta, H. et al. Oligodendrocyte death in Pelizaeus–Merzbacher disease is rescued by iron chelation. *Cell Stem Cell* **25**, 531–541 (2019).
13. Gruenenfelder, F. I. et al. Neural stem cells restore myelin in a demyelinating model of Pelizaeus–Merzbacher disease. *Brain* **143**, 1383–1399 (2020).
14. Elitt, M. S. et al. Chemical screening identifies enhancers of mutant oligodendrocyte survival and unmasks a distinct pathological phase in Pelizaeus–Merzbacher Disease. Stem *Cell Rep*. **11**, 711–726 (2018).
15. Southwood, C. M., Garbern, J., Jiang, W. & Gow, A. The unfolded protein response modulates disease severity in Pelizaeus–Merzbacher disease. *Neuron* **36**, 585–596 (2002).
16. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
17. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
18. Tatar, C. L. et al. Increased *Plp1* gene expression leads to massive microglial cell activation and inflammation throughout the brain. *ASN Neuro* **2**, e00043 (2010).
19. Uschkureit, T., Spörkel, O., Büssow, H. & Stoffel, W. *Rumpshaker*-like proteolipid protein (PLP) ratio in a mouse model with unperturbed structural and functional integrity of the myelin sheath and axons in the central nervous system. *Glia* **35**, 63–71 (2001).
20. Finkel, R. S. et al. Nusinersen versus sham control in infantile-onset spinal muscular atrophy. *N. Engl. J. Med*. **377**, 1723–1732 (2017).
21. Kordasiewicz, H. B. et al. Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. *Neuron* **74**, 1031–1044 (2012).
22. Mazur, C. et al. Brain pharmacology of intrathecal antisense oligonucleotides revealed through multimodal imaging. *JCI Insight* **4**, 129240 (2019).
23. Wu, Q. et al. Elevated levels of the chemokine GRO-1 correlate with elevated oligodendrocyte progenitor proliferation in the *jimpy* mutant. *J. Neurosci*. **20**, 2609–2617 (2000).
24. Miller, M. J. et al. Proteolipid protein gene mutation induces altered ventilatory response to hypoxia in the myelin-deficient rat. *J. Neurosci*. **23**, 2265–2273 (2003).
25. Ueda, A. et al. Pelizaeus–Merzbacher disease can be a differential diagnosis in males presenting with severe neonatal respiratory distress and hypotonia. *Hum. Genome Var*. **5**, 18013 (2018).
26. Osorio, M. J. et al. Concise Review: stem cell-based treatment of Pelizaeus–Merzbacher disease. *Stem Cells* **35**, 311–315 (2017).
27. Renier, W. O. et al. Connatal Pelizaeus-Merzbacher disease with congenital stridor in two maternal cousins. *Acta Neuropathol*. **54**, 11–17 (1981).
28. Lee, Y. et al. Oligodendroglia metabolically support axons and contribute to neurodegeneration. *Nature* **487**, 443–448 (2012).
29. Fünfschilling, U. et al. Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity. *Nature* **485**, 517–521 (2012).
30. Freeman, S. A. et al. Acceleration of conduction velocity linked to clustering of nodal components precedes myelination. *Proc. Natl Acad. Sci. USA* **112**, E321–E328 (2015).

Methods

All data was reproduced with biological replicates as indicated. Blinding was used as indicated. No statistical methods were used to predetermine sample size and the experiments were not randomized. P values shown for $P < 0.1$, otherwise not significant.

Mice

All procedures were in accordance with the National Institutes of Health Guidelines for the Care and Use of Laboratory Animals and were approved by the Case Western Reserve University Institutional Animal Care and Use Committee (IACUC).

Wild-type (B6CBACa-Aw-J/A) and *jimpy* (B6CBACa-Aw-J/A-Plp1jp EdaTa/J; RRID:IMSR_JAX:000287) mice used in this study were purchased from Jackson Laboratory. *jimpy* males possess a point mutation in the splice acceptor site of *Plp1* intron 4 (c.623-2A>G), which results in exclusion of exon 5 and a frameshift of the final 70 amino acids of PLP[31]. The colony was maintained by breeding heterozygous females, which lack a phenotype, to wild-type males to generate affected *jimpy* males. Mice were housed under a temperature-controlled environment, 12 h:12 h light:dark cycle with ad libitum access to water and rodent chow. All mice were genotyped approximately a week after birth using genomic DNA isolated from tail tips or toes at two loci: (1) the *jimpy* mutation (NM_011123.4:c.623-2A>G) in *Plp1* intron 4, which causes skipping of exon 5 and a truncated PLP protein, and (2) the complex indel in *Plp1* exon 3 from dual cutting of CRISPR–Cas9 sgRNAs in CR-*impy* mice (c.[242_318del; 328_330del]). This causes a frameshift in *Plp1*, a premature stop codon in exon 4, and is predicted to cause nonsense-mediated decay of the transcript and loss of protein. Genotyping was performed by standard Sanger sequencing or custom real-time PCR assays (probe identifiers: Plp1-2 Mut [for *jimpy* mutation in intron 4] and Plp1-5 WT [for CR-*impy* complex deletion in exon 3], Transnetyx).

Primers for Sanger sequencing are provided in Supplementary Table 4.

Design of *Plp1*-targeting sgRNA

Mouse *Plp1* sequence was entered into the *Streptococcus pyogenes* CRISPR–spCas9 sgRNA design tool at crispr.mit.edu[32] and analysed against the mm10 target genome. *Plp1*-targeting sgRNAs were sorted on the basis of their on-target efficiency while minimizing off-target mutations. On-target nuclease activity was confirmed for each sgRNA using the Guide-it sgRNA Screening Kit (631440, Clontech) according to the manufacturer's instructions. The following sgRNAs were tested: sgRNA1, CCCCTGTTACCGTTGCGCTC; sgRNA2, TGGCCACCA GGGAAGCAAAG; sgRNA3, AAGACCACCATCTGCGGCAA; sgRNA4, GGCCTGAGCGCAACGGTAAC; sgRNA5, GCCTGAGCGCAACGGTAACA; sgRNA6, TCTACACCACCGGCGCAGTC; sgRNA7, CCAGCAGGAGGGCCC CATAA; and sgRNA8, GAAGGCAATAGACTGACAGG.

This list was further filtered on the basis of the ability of each sgRNA to target *Plp1*'s splice isoform *Dm20*, in addition to *Plp1*. We selected two sgRNAs (3 and 7) that targeted exon 3 of *Plp1* for combined use in zygote studies, which enabled the rapid detection of large deletion events by PCR and provided redundancy for on-target cutting.

Suppression of *Plp1* in *jimpy* zygotes using CRISPR–Cas9

Carrier female oocyte donors were administered 5 IU pregnant mare serum gonadotropin by intraperitoneal injection (G4877, Sigma-Aldrich), followed by 2.5 IU human chorionic gonadotropin (GC10, Sigma-Aldrich) 48 h later. These superovulated females were mated to wild-type males. Zygotes were collected in FHM medium (MR-025 Sigma-Aldrich) with 0.1% hyaluronidase (H3501, Sigma-Aldrich) and the surrounding cumulus cells were separated. The zona pellucida of each zygote was partially dissected using 0.3 M sucrose (S7903, Sigma-Aldrich) in FHM as previously described[33].

Zygotes were placed in 2× KSOM medium (MR-106, Sigma-Aldrich) with an equal volume of solution containing 100 ng μl^{-1} sgRNA3, 100 ng μl^{-1} sgRNA7 (AR01, PNAbio), and 200 ng μl^{-1} spCas9 mRNA (CR01, PNAbio). Given the low frequency of *jimpy* zygotes and unknown in vivo targeting of the sgRNAs, both sgRNAs were used simultaneously to maximize the chance of *Plp1* frameshift. Electroporation was performed in a chamber with a 1-mm gap between two electrodes using an ECM 830 Square Wave Electroporation System (BTX). Electroporation parameters were set as follows: 32 V, 3 ms pulse duration, 5 repeats and 100 ms inter-pulse interval. Electroporated zygotes were moved to KSOM medium and then transferred into the oviducts of pseudo-pregnant females (CD1). Electroporation settings were optimized to achieve maximal cutting efficiency in a separate strain but resulted in a higher rate of embryo loss in our B6CBACa/J strain. Zygotes were electroporated in batches of 54, 56 and 61, which resulted in 4, 3 and 0 pups born. The seven surviving mice were genotyped after birth and monitored daily for onset of typical *jimpy* phenotypes including tremors, seizures and early death by postnatal day 21. A founder *jimpy* male with complex deletion containing 80 bp of total deleted sequence in exon 3 of *Plp1*, denoted CR-*impy*, showed no overt phenotype and was back-crossed for two generations to the wild-type parental strain to reduce potential off-target Cas9 cutting effects (Extended Data Fig. 1b–e). A colony of mice was bred to evaluate cellular, molecular, and functional phenotypes of contemporaneous isogenic wild-type, *jimpy* and CR-*impy* male mice. Mice were monitored daily to determine lifespan with statistical significance among groups determined using the log-rank test. Additionally, mice surviving beyond three weeks were analysed using behavioural (rotarod and open-field testing for motor performance), histology (immunostaining of the CNS for myelin proteins and electron microscopy for myelin ultrastructure) and electrophysiology (conduction velocity of the optic nerve). Details and metadata for all mice in this study including censoring of animals in the survival analysis are found in Supplementary Data 1.

CRISPR on- and off-target assessment

CRISPR on-target cutting efficiencies were assessed by high-throughput sequencing. PCR primers were designed to encompass each guide on-target site. Primer sequences were generated using NCBI Primer-BLAST and are provided in Supplementary Table 4. The following tails were added to the primer sequences: forward, TCCCT ACACGACGCTCTTCCGATCT; and reverse, AGTTCAGACGTGTGCTCT TCCGATCT.

PCR amplification on tail-tip genomic DNA was performed using the KAPA HiFi HotStart ReadyMix (07958935001, Roche) to minimize PCR-based error. Libraries were prepared by adding unique indices by PCR using KAPA HiFi HotStart ReadyMix. All libraries were pooled evenly and quantified using NEBNext Library Quant Kit for Illumina (E7630, New England Biolabs) then denatured and diluted per Illumina's MiSeq instructions. Then, 250 bp paired-end sequencing was performed using an Illumina MiSeq at the Case Western Reserve University School of Medicine Genomics Core Facility. Reads were compared against the consensus sequence and CRISPR-induced indel percentages were determined using the OutKnocker tool[34] (http://outknocker.org).

Genomic DNA was isolated from brain tissue from the CR-*impy* founder male, three F_2 generation CR-*impy* male mice (each from a unique breeding pair using independent F_1 generation carrier females), and a *jimpy* male from a contemporaneous but independent cohort in our colony. Libraries were prepared for whole genome sequencing using Nextera DNA Flex Library prep (20018705, Illumina) and 150 bp paired-end sequencing was performed using an Illumina NovaSeq. Reads were aligned to the mouse genome (mm10) using BWA[35] (v.0.7.17-r1188) with default parameters for paired reads. Local indel realignment was performed using GATK RealignerTargetCreator and IndelRealigner (v.3.3-2-gec30cee) at the on-target and off-target sites.

Reads aligned to the window chrX:136831817-136832360 at the *Plp1* locus were re-aligned using Blat (v.36x2) to fully capture the CR-*impy* complex deletion.

The top-50 potential off-target sites for each sgRNA were identified using the CCTop - CRISPR–Cas9 target online predictor tool[36], with a maximum total mismatch number of 4. Additionally, each site was identified using the RGEN Cas-OFFinder[37] and CRISPOR[38] off-target prediction algorithms, providing two independent validations of this off-target location list. The indel-realigned reads were visually inspected in Integrative Genomics Viewer (IGV)[39], and indels occurring at a frequency of at least 5% after filtering known polymorphisms from dbSNP (build 142) at these 50 potential off-target sites were considered CRISPR-induced mutations.

Video recording of mouse phenotypes

All recording was performed using video recording function on an Apple iPhone. Videos were colour corrected, stabilized and trimmed to a discrete range using Apple iMovie. Videos were collated and converted to MP4 format using Adobe After Effects.

Immunohistochemistry

Mice were anaesthetized with isoflurane and euthanized by transcardial perfusion with PBS followed by 4% paraformaldehyde (PFA). Tissue was collected and placed in 4% PFA overnight at 4 °C. Samples were rinsed with PBS, equilibrated in 30% sucrose, and frozen in Tissue-Tek Optimum Cutting Temperature compound (OCT; 25608-930, VWR). Samples were cryosectioned at a 20 μm thickness. Sections were washed in phosphate-buffered saline (PBS) and incubated overnight in antibody solution containing 2.5% normal donkey serum (NDS; 017-000-121, Jackson Laboratories) and 0.25% Triton X-100 (T8787, Sigma).

Alternatively, as noted elsewhere in the Methods, mice were euthanized by CO_2 asphyxiation, followed by tissue collection, immersion fixation overnight in 10% neutral-buffered formalin, and paraffin embedding. Sections 5 μm thick were cut onto charged glass slides and dried overnight at 60 °C. Sections were deparaffinized and hydrated using graded concentrations of ethanol to deionized water. Sections were subjected to antigen retrieval by sodium citrate buffer at pH 6 (H-3300; Vector Laboratories) at 100 °C for 45 min, gently washed in deionized water, and then transferred into 0.05 M Tris-based solution in 0.15 M NaCl with 0.1% (v/v) Triton X-100, pH 7.6 (TBST). For chromagen staining, endogenous peroxidase was blocked with 3% hydrogen peroxide for 20 min. Nonspecific background staining was blocked in 3% normal goat serum for 30 min (Sigma) at room temperature. For mouse antibodies, sections were incubated for 30 min in Mouse Blocking Reagent (Vector Laboratories). All slides were then incubated at 4 °C overnight with cocktails of primary antibodies in TBST. For DAB reactions, after washing with TBST, sections were then incubated with the species-appropriate immunoglobulin G (IgG)-horseradish peroxidase (HRP) (1:300, SC2004; Santa Cruz), then reacted with diaminobenzidine (DAB; ScyTek Laboratories) and counterstained with haematoxylin (no. 7211; Richard-Allen Scientific).

Sections were stained using the following antibodies at the indicated concentrations or dilutions: mouse anti-MBP (2 μg ml^{-1}; 808401, Biolegend; RRID:AB_2564741), rabbit anti-MBP (1:1,000; Abcam, ab40390; RRID:AB_1141521), rabbit anti-MyRF polyclonal antibody (1:500; provided by M. Wegner), goat anti-SOX10 (0.4 μg ml^{-1}; AF2864, R&D Systems; RRID:AB_442208), rabbit anti-GFAP (1:1,000; Z0334, Dako; RRID:AB_10013382), goat anti-IBA1 (0.1 mg ml^{-1}; ab5076, Abcam), rabbit anti-IBA1 (1:2,000; 019–19741, WAKO; RRID:AB_839504), rabbit anti-ASO (1:2,500; Ionis Pharmaceuticals), rabbit anti-HDAC2 (1:250; Abcam, ab16032; RRID:AB_2118543), mouse anti-APC/CC1 (2.5 μg ml^{-1}; ab16794, Abcam; RRID:AB_443473), mouse anti-APC/CC1 (1:250; MABC200, Millipore; RRID:AB_11203645), rat anti-NG2 (25 μg ml^{-1}; MAB6689, R&D Systems; RRID:AB_10890940), goat anti-PDGFRα (1:500; AF1062, R&D systems; RRID:AB_2236897) and rabbit anti-OLIG2 (1:250; 13999-1-AP, ProteinTech; RRID:AB_2157541). For MBP immunohistochemistry, sections were post fixed in methanol at -20 °C for 20 min followed by overnight incubation in a PBS based primary antibody solution containing 0.1% saponin and 2.5% normal donkey serum. Secondary immunostaining was performed with Alexa Fluor antibodies (ThermoFisher) used at 1 μg ml^{-1}. Nuclei were identified using DAPI (100 ng ml^{-1}; D8417, Sigma). Stained sections were imaged using the Operetta High Content Imaging and Analysis system (PerkinElmer) and Harmony software (PerkinElmer) for whole-section images and a NanoZoomer S60 Digital slide scanner (Hamamatsu) for all other immunohistochemical imaging, unless otherwise noted.

To quantify MyRF, SOX10, OLIG2, CC1 or PDGFα-positive cells, counts were performed along the length of the whole corpus callosum, the cerebellum and the brainstem in medial sagittal sections from three animals per genotype. CC1 and OLIG2 or PDGFα and OLIG2 double-positive were determined from these counts. Counts were performed in a semi-automated manner using ImageJ (National Institutes of Health). To quantify GFAP and IBA1 staining, fluorescence intensity was measured using Adobe Photoshop along the length of the whole corpus callosum, the cerebellum and the brainstem from medial sagittal sections from three animals per genotype. To quantify cleaved caspase 3 staining, sections from regions starting at the sagittal midline to 600 μm from the midline were used and cleaved caspase 3 positive cells were counted along the entire length of the corpus callosum, white matter of the cerebellum and entire brainstem to determine the total number of apoptotic cells per treatment group. All counts and quantifications were performed in a blinded manner. One-way ANOVA with Tukey's correction and two-way unpaired *t*-tests, or a one-way ANOVA with Dunnett's correction for multiple comparisons were used to determine statistical significance across CRISPR or ASO cohorts, respectively.

RT–qPCR

Mice from CRISPR or ASO studies were euthanized using isoflurane overdose. Different brain regions (cerebral cortex, cerebellum and brainstem) were collected and flash frozen. Each region was split in two and half was used for RNA quantification using RT–qPCR, the other for western blot analysis (see below). TRI Reagent (R2050-1-200, Zymo Research) was separately added to tissue and samples were homogenized using Kontes Pellet Pestle Grinders (KT749520-0000, VWR). RNA was extracted using the RNeasy Mini Kit (74104, Qiagen) according to the manufacturer's instructions. Reverse transcription was performed using the iScript cDNA Synthesis Kit (1708891, Biorad) with 1 μg of RNA per reaction. Real-Time PCR was then performed on an Applied Biosystems 7300 Real-time PCR system with 10 ng cDNA per sample in quadruplicate using Taqman gene expression master mix (4369016, ThermoFisher) and the following pre-designed Taqman gene expression assays (4351370, ThermoFisher): *Plp1* (Mm01297210_m1), *Mbp* (Mm01266402_m1) and *Actb* (Mm00607939_s1) (endogenous control). Expression values were normalized to *Actb* and to wild-type samples (for CRISPR cohort) or wild-type untreated samples (for ASO-treated wild-type cohort). One-way ANOVA with Tukey's correction and two-way unpaired *t*-tests, or a one-way ANOVA with Dunnett's correction for multiple comparisons were used to determine statistical significance across CRISPR or ASO cohorts, respectively.

Protein quantification and western blot

Tissues were obtained as described above. Protein lysis buffer consisting of RIPA buffer (R0278, Sigma), cOmplete Mini EDTA-free Protease Inhibitor Cocktail (11836170001, Sigma), Phosphatase Inhibitor Cocktail 3 (P0044, Sigma), Phosphatase Inhibitor Cocktail 2 (P5726, Sigma), and BGP-15 (B4813, Sigma) was added to each sample. Tissue was homogenized using Dounce Tissue Grinders (D8938, Sigma). Lysate was separated by centrifugation at 17,000*g* for 15 min at 4 °C. A BCA standard curve was generated using the Pierce BCA Protein Assay Kit (23225, Thermo Scientific) and used to samples to an equivalent protein concentration. Equal amounts of sample were run on a NuPAGE 4–12% Bis-Tris Protein gel (NP0335BOX or NP0329BOX, Thermo Fisher), then electrophoretically transferred to a PVDF membrane (LC2002, Invitrogen or 926-31097, Li-Cor). The membrane was blocked with 5% milk in TBS-T for an hour, then hybridized with mouse anti-MBP antibody (1 μg ml^{-1}; 808401, Biolegend; RRID:AB_2564741) or rat anti-PLP antibody (1:1,000; clone AA3, Lerner Research Institute Hybridoma Core) overnight at 4 °C. Blots were then washed in TBS-T and incubated in goat anti-mouse HRP (1:2500, 7076, Cell Signaling), goat anti-rat HRP (1:2500, 7077, Cell Signaling) or IRDye secondary antibodies (1:20,000, 925, Li-Cor). Each sample was normalized to β-actin using HRP-conjugated mouse anti-β-actin (1:10,000, A3854, Sigma-Aldrich; RRID:AB_262011). All secondary antibodies were incubated for one hour at room temperature. Blots were analysed with the Odyssey Fc imaging system (Li-Cor). One-way ANOVA with Tukey correction and two-way unpaired *t*-tests, or a one-way ANOVA with Dunnett's correction for multiple comparisons were used to determine statistical significance across CRISPR or ASO cohorts, respectively. Raw annotated images of full western blots are provided in Supplementary Data 2, 7.

Sample preparation for label-free expression discovery

Samples in protein lysis buffer were cleaned of detergent as described[40], with a 10-kDa molecular weight cutoff filter (Millipore) and buffer exchanged with 8 M urea in 50mM Tris pH 8.0 to a final volume of 50 μl. Proteins were reduced on filter with 10 mM dithiothreitol (8 M urea, 50 mM Tris-pH-8.0) for 1 h at 37 °C, followed by alkylation with 25 mM iodoacetaminde (8 M urea, 50 mM Tris pH 8.0) for 30 min in the dark. The 8 M urea was then adjusted to 4 M (50 mM Tris pH 8.0) and samples were concentrated to a final volume of 50 μl. Next, 10 μg of total protein were digested with lysyl endopeptidase (Wako Chemicals) at an enzyme:substrate ratio of 1:30 for 2 h at 37 °C. The urea concentration was then adjusted to 2 M using 50 mM Tris, pH 8, followed by an overnight trypsin digestion using sequencing grade trypsin (Promega) at an enzyme:substrate ratio of 1:30 at 37 °C.

Reverse phase LC–MS/MS analysis

Three hundred nanograms of each sample were analysed by LC–MS/MS using a LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific) equipped with a nanoAcquity Ultra-high pressure liquid chromatography system (Waters). The injection order on the LC–MS was randomized over all samples. Blank injections were run after each sample to minimize carry-over between samples. Mobile phases were organic phase A (0.1% formic acid in water) and aqueous phase B (0.1% formic acid in acetonitrile). Peptides were loaded onto a nanoACQUITY UPLC 2G-V/M C18 desalting trap column (180 μm × 20 mm nano column, 5 μm, 100 Å) at flow rate of 0.300 μl min^{-1}. Subsequently, peptides were resolved in a nanoACQUITY UPLC BEH300 C18 reversed-phase column (75μm × 250 mm nano column, 1.7 μm, 100 Å; Waters) followed by a gradient elution of 1–40% of phase B over 240 min (isocratic at 1% B, 0–1 min; 2–42% B, 2–212 min; 42–90% B, 212–223 min; and 90–1% B, 223–240 min). A nano ES ion source at a flow rate of 300 nl min^{-1}, 1.5 kV spray voltage and 270 °C capillary temperature was used to ionize peptides. Full scan MS spectra (*m*/*z* 380–1,800) were acquired at a resolution of 60,000 followed by twenty data dependent MS/MS scans. LC–MS/MS raw data were acquired using the Xcalibur software (Thermo Fisher Scientific, v.2.2 SP1).

Data processing for protein identification and quantification

The LC–MS/MS raw files (one for each sample) were imported into PeaksStudio (BioinformaticsSolutions) and processed as previously described[41,42]. A database was created that included PLP wild-type and predicted mutant isoforms. Search settings were as follows: trypsin enzyme specificity; mass accuracy window for precursor ion, 10 ppm; mass accuracy window for fragment ions, 0.8 Da; carbamidomethylation of cysteines as fixed modifications; oxidation of methionine as variable modification; and one missed cleavage. Peptide identification criteria were a mass accuracy of ≤ 10 ppm, and an estimated false discovery rate of less than 2%. Normalization of signal intensities across samples was performed using the average signal intensities obtained in each sample. The fold change was then calculated using these average intensity values for the protein across the two samples.

Electron microscopy

Mice were anaesthetized with isoflurane and tissue was collected after terminal transcardial perfusion with PBS followed by 4% paraformaldehyde and 2% glutaraldehyde (16216, Electron Microscopy Sciences) in 0.1M sodium cacodylate buffer, pH 7.4 (11652, Electron Microscopy Sciences), except for 6-month optic nerve samples which were placed directly into fixative without perfusion. Samples were post-fixed with 1% osmium tetroxide (19150, Electron Microscopy Sciences) and stained with 0.25% uranyl acetate (22400, Electron Microscopy Sciences) en bloc. Samples were dehydrated using increasing concentrations of ethanol, passed through propylene oxide, and embedded in Eponate 12 epoxy resin (18012, Ted Pella). Silver–coloured sections were prepared (Leica EM UC6), placed on 300 mesh nickel grids (T300-Ni, Electron Microscopy Sciences), stained with 2% uranyl acetate in 50% methanol, and stained with lead citrate (17800, Electron Microscopy Sciences). Sections were imaged using a FEI Tecnai Spirit electron microscope at 80 kV. Myelinated axons were manually counted from the sections made on the middle portion of the optic nerve lengthwise, the medial portion of the genu for the corpus collosum, and corticospinal tracts at the pontine level of the brainstem. Three independent areas were counted for each region using Adobe Photoshop (Adobe Systems). Two-way unpaired *t*-tests or a one-way ANOVA with Dunnett's correction for multiple comparisons were used to determine statistical significance across CRISPR or ASO cohorts, respectively.

Optic nerve electrophysiology

Mice were deeply anaesthetized with isoflurane and euthanized. Each eye with its attached optic nerve was dissected and placed in Tyrode's solution consisting of 129 mM NaCl (BP358-212, Fisher Scientific), 3 mM KCl (BP366-500, Fisher Scientific), 1.2 mM NaH_2PO_4 (1-3818, J. T. Baker Chemical), 2.4 mM $CaCl_2$ (C79-500, Fisher Scientific), 1.3 mM $MgSO_4$ (M2643, Sigma), 20 mM $NaHCO_3$ (S233-500, Fisher Scientific), 3 mM HEPES (H3375, Sigma), 10 mM glucose (G5767, Sigma), oxygenated using a 95% O_2/5% CO_2 gas mixture. Each nerve was carefully cleaned, transected behind the eye at the optic chiasm, and allowed to recover for 1 h in oxygenated Tyrode's solution at room temperature (22–24 °C). Each end of the nerve was set in suction electrodes, pulled from polyethylene tubing (PE-190, BD Biosciences). Monophasic electrical stimuli were applied to the proximal end of the nerve and recordings were captured at the distal end. The recovery of the response was monitored every 20 min for 1 h, and only fully recovered samples were subjected to additional stimuli. Stimuli were generated with a S48 stimulator (Grass Technologies) and isolated from ground with PSIU6B unit (Grass Technologies). Supra-threshold stimulus was determined using 30-μs stimulus duration. The response was amplified 100× with a P15D preamplifier (Grass Technologies), monitored with oscilloscope (V1585, Hitachi), digitized with Digidata1550A (Axon Instruments) and recorded using 50-kHz sampling rate with AxoScope software (Axon Instruments). The distance between the electrodes was measured and used to calculate the conduction velocity of the compound action potential peaks at their latency. Recorded signals were analysed using AxoScope software. One-way ANOVA with Tukey correction and two-way unpaired *t*-tests, or a one-way *t*-test were used to determine statistical significance across CRISPR or ASO cohorts, respectively.

Open-field testing

Locomotion was assessed by open-field testing. Animals were placed in the centre of a 20-inch by 20-inch square box and all movements were captured for a total of 5 min using ANY-maze software v.5.0

(Stoelting). Total distance travelled was reported for each animal. One-way ANOVA with Tukey correction and two-way unpaired *t*-tests, or a one-way ANOVA with Dunnett's correction for multiple comparisons were used to determine statistical significance across CRISPR or ASO cohorts, respectively.

Rotarod testing

Motor performance was assessed using a Rota Rod Rotomax 5 (Columbus Instruments) with a 3-cm diameter rotating rod. Immediately before testing, animals were trained at a constant speed of 4 rounds per minute (rpm) for a total of 2 min. Testing began at 4 rpm with an acceleration of 0.1 rpm s^{-1}. Time to fall was recorded from three independent trials, and the average value for each animal was reported. Animals were allowed to rest for at least 5 min between training and each experimental trial. Animals that failed training were assigned a value of 0 for all three trials for a particular time point. One-way ANOVA with Tukey correction and two-way unpaired *t*-tests, or a one-way ANOVA with Dunnett's correction for multiple comparisons were used to determine statistical significance across CRISPR or ASO cohorts, respectively.

Immunocytochemistry

Cells were fixed with 4% PFA in PBS. After fixation, cells were permeabilized with 0.2% Triton X-100 in PBS followed by blocking in 10% donkey serum in PBS. Cells were stained overnight at 4 °C with the following primary antibodies diluted in blocking solution: mouse anti-MBP (1:500; 808401, Biolegend; RRID:AB_2564741), rat anti-PLP (1:5,000; clone AA3, Lerner Research Institute Hybridoma Core), goat anti-SOX10 (2 μg ml^{-1}; AF2864, R&D Systems; RRID:AB_442208), rabbit anti-OLIG2 (1:1,000; 13999-1-AP, ProteinTech; RRID:AB_2157541), rabbit anti-NANOG (0.4 μg ml^{-1}; AB21624, Abcam; RRID:AB_446437), mouse anti-OCT3/4 (0.4 μg ml^{-1}; SC-5279, Santa Cruz; RRID:AB_628051). For secondary immunostaining, Alexa Fluor antibodies (ThermoFisher) were used at 1 μg ml^{-1}, and DAPI (100 ng ml^{-1}) was used to identify nuclei. Images were captured using Leica DMi8 fluorescence microscope (induced pluripotent stem (iPS) cells) or Operetta High Content Imaging and Analysis system and Harmony software (OPCs and oligodendroctytes), the latter quantified using Columbus software (PerkinElmer).

Generation of iPS cells

Tail tips (2 mm piece from 8-day-old CR-*impy* mice) were bisected, placed on Nunclon-Δ 12-well plates (150628, ThermoFisher), and covered with a circular glass coverslip (12-545-102; Fisher Scientific) to maintain tissue contact with the plate and enable fibroblast outgrowth. Tail-tip fibroblasts were cultured in fibroblast medium consisting of DMEM (11960069, ThermoFisher) with 10% fetal bovine serum (FBS; 16000044, ThermoFisher), 1× non-essential amino acids (11140050, ThermoFisher), 1× Glutamax (35050061, ThermoFisher) and 0.1 mM 2-mercaptoethanol (M3148, Sigma Aldrich) supplemented with 100 U ml^{-1} penicillin–streptomycin (15070-063, ThermoFisher). Medium was changed every day for the first 3 days and then every other day.

Fibroblasts were seeded at approximately 1.4×10^4 cells per cm^2 on Nunclon-Δ dishes in fibroblast medium, and allowed to equilibrate overnight. The following day, medium was removed and replaced with an equal volume of pHAGE2-TetOminiCMV-STEMCCA-W-loxp lentivirus encoding a floxed, doxycycline-inducible polycistronic Oct4, Sox2, Klf4 and c-Myc construct and pLVX-Tet-On-Puro (632162, Clontech) lentivirus supplemented with 8 μg ml^{-1} polybrene (107689, Sigma). Lentivirus was prepared using the Lenti-X Packaging Single Shots (631275, Clontech) according to manufacturer's instructions. Three hours later lentivirus medium was removed and replaced with fibroblast medium supplemented with 2 μg ml^{-1} doxycycline (631311, Clontech). The following day, medium was removed and replaced with an equal volume of pHAGE2-TetOminiCMV-STEMCCA-Wloxp and pLVX-Tet-On-Puro lentivirus supplemented with 8 μg ml^{-1} polybrene. Three hours later lentivirus medium was diluted 1:2 with fibroblast medium. Medium was changed each day with fibroblast medium supplemented with 2 μg ml^{-1} doxycycline and 10^3 units per ml LIF. After 3 days, fibroblasts were lifted using Accutase and seeded on Nunclon-Δ plates, on a feeder layer of irradiated mouse embryonic fibroblasts (iMEFs; produced in-house) previously plated at 1.7×10^4 cells per cm^2 on 0.1% gelatin (1890, Sigma) coated Nunclon-Δ plates in pluripotency medium consisting of Knockout DMEM (10829-018, ThermoFisher), 5% FBS, 15% knockout replacement serum (10828028, ThermoFisher), 1× Glutamax, 1× nonessential amino acids, 0.1 mM 2-mercaptoethanol, and 10^3 units per ml LIF (LIF; ESG1107, EMD Millipore) supplemented with 2 μg ml^{-1} doxycycline. Medium was changed every day until iPS cell colonies began to emerge. Individual colonies were picked and dissociated in Accutase and were individually plated in single wells of Nunclon-Δ 12-well plates, atop an iMEF feeder layer in pluripotency medium supplemented with 2 μg ml^{-1} doxycycline. Clones were further expanded, with daily medium changes. iPS cell colonies were stained for pluripotency markers Nanog and Oct4 and karyotyped at the seventh passage after derivation (Cell Line Genetics). CR-*impy* iPS cells were derived and characterized for this study (line identifier jpCR100.1). Isogenic comparator *jimpy* (line identifier i.jp-1.6) and wild-type (line identifier i.wt-1.0) iPS cell lines were described and characterized separately[14]. All cell cultures in the laboratory are routinely tested for mycoplasma contamination with consistently negative results. Genotypes of iPS cells were re-verified before use. For characterization iPS cells were immunostained for Nanog and OCT3/4, and counterstained with DAPI.

Generation of iPS-cell-derived OPCs

iPS cells were differentiated to OPCs as previously described[43,44]. In brief, iPS cells were isolated from their iMEF feeder layer using 1.5 mg ml^{-1} collagenase type IV (17104019, ThermoFisher) and dissociated with either 0.25% Typsin-EDTA or Accutase and seeded at 7.8×10^4 cells per cm^2 on Costar Ultra-Low attachment 6-well plates (3471, Corning). Cultures were then directed through a stepwise differentiation process to generate pure populations of OPCs. OPCs were maintained in OPC medium consisting of DMEM/F12 (11320082, ThermoFisher), 1× N2 supplement (AR009, R&D Systems), 1× B-27 without vitamin A supplement (12587-010, ThermoFisher), and 1× Glutamax (collectively N2B27 medium), supplemented with 20 ng ml^{-1} fibroblast growth factor 2 (FGF2; 233-FB, R&D Systems) and 20 ng ml^{-1} platelet-derived growth factor-AA (PDGF-AA; 221-AA, R&D Systems). Medium was changed every other day. All cell cultures in the laboratory are routinely tested for mycoplasma contamination with consistently negative results. For characterization of purity, iPS-cell-derived OPCs were fixed with 4% PFA and immunostained for canonical OPC transcription factors, OLIG2 and SOX10, and counterstained with DAPI.

In vitro assessment of oligodendrocyte differentiation from OPCs

OPCs from each genotype were plated in parallel onto Nunclon-Δ 96-well plates (150628, ThermoFisher) that were first coated with 100 μg ml^{-1} poly(L-ornithine) (P3655, Sigma), followed by 10 μg ml^{-1} laminin solution (L2020, Sigma). For the oligodendrocyte differentiation assay, 25,000 cells were seeded per well in medium that consisted of DMEM/F12 (11320082, ThermoFisher), 1× N2 supplement (AR009, R&D Systems), 1× B-27 without vitamin A supplement (12587-010, ThermoFisher) and 1× Glutamax, supplemented with T3 (40 ng ml^{-1}), Noggin (100 ng ml^{-1}), cAMP (10 μM), IGF (100 ng ml^{-1}) and NT3 (10 ng ml^{-1}). All plates were incubated at 37 °C and 5% CO_2 for 3 days. Cells were fixed and immunostained for MBP and PLP, and counterstained with DAPI. All quantifications were normalized to initial cell counts at plating.

Assessment of gene expression modulation in the oligodendrocyte lineage by *Hdac2*-targeting ASOs

Two ASOs were designed to target mouse *Hdac2*. ASO-*Hdac2*.a consisted of a 20-mer nucleotide sequence (5′-CTCACTTTTCGAGGTTCCTA-3′)

with 2′-O-methoxyethyl modifications and a mixed backbone of phosphorothioate and phosphodiester internucelotide linkages. ASO-*Hdac2*.b consisted of a 16-mer nucleotide sequence (5′-CATCATCTATACCATC-3′) with 2′-O-ethyl modifications with a full backbone of phosphorothioate internucelotide linkages. To determine whether ASOs could reduce effectively target oligodendrocyte linage cells and reduce gene expression, we administered *Hdac2*-targeting ASOs to 8-week-old C57BL/6J mice (Jackson Labs) via single 300 µg ICV injection. After 2 weeks, mice were euthanized and processed for histology. Formalin-fixed, paraffin embedded brain and spinal cord sections were stained for NG2 to label OPCs in the study dosed with *Hdac2*.a ASO, and APC/CC1 to oligodendrocytes in the study dosed with *Hdac2*.b ASO, as well for HDAC2 to examine ASO-mediated knockdown of this target. Images were captured using an epifluorescent imaging system (EVOS, ThermoFisher Scientific).

Plp1-targeting ASO design and characterization

Second generation ASOs were designed to target mouse *Plp1*. ASOs consisted of 20-mer nucleotide sequences with 2′-O-methoxyethyl modifications and a mixed backbone of phosphorothioate and phosphodiester internucelotide linkages. ASOs were screened for efficacy in primary E16 cortical cultures, as previously described[45]. In brief, cells were treated with ASOs at 37 °C/5% CO_2 for 3 days, RNA was isolated, and *Plp1* transcript level was quantified with RT–qPCR on Step One instruments (Thermo Fisher). *Plp1* mRNA was normalized to total RNA measured with the Quant-iTTM RiboGreen RNA reagent. ASOs that efficiently reduced *Plpl* mRNA were selected for in vivo screening and tolerability studies.

Lead ASOs were administered to 8-week-old C57BL/6J mice via single 500 µg ICV injection and *Plp1* mRNA levels were measured by RT–qPCR in cortex and spinal cord tissue after 2 weeks. ASOs with greater than 90% *Plp1* mRNA reduction were selected for further characterization. Selected ASOs were administered to mice via single 300 µg ICV bolus injection to test for efficacy and tolerability, as measured by markers of glial cell activation, 8 weeks after ICV injection. Levels of *Plp1* mRNA as well as markers of astrocytes, microglia, and monocytes (*Gfap*, *Aif1* and *Cd68*, respectively) were assessed by RT–qPCR using the custom primer and probe sets (Integrated DNA Technologies) listed in Supplementary Table 4.

Immunohistochemical staining was used to assess morphology of astrocytes, microglia, and oligodendrocyte using anti-GFAP, IBA1 (DAKO), and MBP (Abcam) antibodies, respectively, in formalin-fixed, paraffin embedded brain and spinal cord sections. *Plp1* ASO.a (intron 5) and ASO.b (3′ UTR) were selected for use in *jimpy* mice, as well as a control ASO with no known murine target. ASO sequences were as follows: ASO control, 5′-CCTATAGGACTATCCAGGAA-3′; ASO *Plp1*.a, 5′-GCTCATTGATTCAAGTACAT-3′; and ASO *Plp1*.b, 5′-GCATTTACCCGAAGGCCATT-3′.

Each *Plp1*-targeting ASO was further evaluated for potential off-target effects. Bowtie aligner[46] was used to identify putative ASO off-target transcript sequences, with up to three base mismatches. This analysis identified potential off-target sequence in *Xylt1* for ASO *Plp1*.a and *Scfd1* and *Tpk1* for ASO *Plp1*.b, each having exactly two mismatches. To determine whether these transcripts were targeted by *Plp1* ASO.a or ASO.b, adult mice (8 weeks of age, C57Bl6/J) were administered 30, 100 or 300 µg of each ASO by ICV injection. After two weeks, spinal cord tissues were collected and levels of *Xylt1*, *Scfd1* and *Tpk1* were measured by RT–qPCR using the custom primer and probe sets (Integrated DNA Technologies) listed in Supplementary Table 4.

Optimum therapeutic dosage for use in early postnatal injection was determined by injecting wild-type C57BL/6J mice pups at postnatal day 1 using three different doses (10, 30 or 60 µg) of ASO *Plp1*.a or ASO *Plp1*.b, along with a control non-targeting ASO. Mice were euthanized three weeks later and analysed by for levels of *Plp1* mRNA in the spinal cord using RT–qPCR. One-way ANOVA with Dunnett's correction for multiple comparisons was used to determine statistical significance across treatments.

Therapeutic application of ASOs to postnatal mice

Male pups from crosses between *jimpy* mutation carrier females and wild-type males were administered 30 µg of either *Plp1*-targeting ASOs *Plp1*.a, *Plp1*.b, a control non-targeting ASO, or left untreated. ASOs were administered using a Hamilton 1700 gastight syringe (7653-01, Hamilton Company) by ICV injection to cryoanaesthetized mice. The needle was placed between bregma and the eye, 2/5 the distance from bregma, and inserted to a depth of 2 mm (ref. [47]). A total volume of 2 µl was administered to the left ventricle. Mice were allowed to recover on a heating pad and subsequently reintroduced to the dam. Injections were performed with the investigator blinded to the genotype.

Mice were genotyped during the first postnatal week and monitored daily for onset of typical *jimpy* phenotypes including tremors, seizures and early death by 3 weeks of age. Lifespan was determined for each animal with statistical significance among groups determined using the log-rank test. All mice surviving to a pre-determined end point of 8 months of age were euthanized for histological analysis. Additionally, animals were analysed using rotarod, open-field and optic nerve electrophysiology. Details and metadata for all mice in this study are found in Supplementary Data 6.

Evaluation of respiration

At postnatal day 19 or 20, male pups were placed in a plethysmograph chamber and pressure changes caused by animal respiration were measured using a differential pressure transducer (Emka). The data collection was started when the mice were placed in the chamber and continuously recorded at 1 kHz sampling rate. After placing the mice in the chamber, it was first flushed with normal air (79% nitrogen, 21% oxygen) over a 1 h period to acclimatize the mice and determine basal breathing activity. The chamber was then flushed with hypercapnic gas (74% nitrogen, 21% oxygen, 5% carbon dioxide) for 15 min and the data collected over the subsequent 15–30 min period were used for analysis. Next, the chamber was flushed with normal air for 15 min. Hypoxic gas (89.5% nitrogen, 10.5% oxygen) was then introduced to the chamber over 10 min, with the data collected over this period used for analysis. After the hypoxic gas challenge, mice were weighed and euthanized. Gas flow rate over the entire experiment was 0.75 l min^{-1} per chamber. Recorded breaths lasting for at least 20 s, continuously, and marked with a 100% success rate using IOX2 software (Emka) were used for subsequent data analysis for the normal air and hypercapnic conditions. Recorded breaths in the hypoxic condition were not continuous for more than 20 s so only breaths marked with a 100% success rate in the IOX2 software were used for further data analysis. Survival during hypoxic challenge was determined for each animal with statistical significance among groups determined using the log-rank test. Variability of respiration was determined with statistical significance among groups determined using the Brown and Forsythe's test.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data generated or analysed during this study are included in this article and its Supplementary Information. Animals and iPS cell lines are available from P.J.T. upon request. Source data are provided with this paper.

31. Nave, K. A., Lai, C., Bloom, F. E. & Milner, R. J. Jimpy mutant mouse: a 74-base deletion in the mRNA for myelin proteolipid protein and evidence for a primary defect in RNA splicing. *Proc. Natl Acad. Sci. USA* **83**, 9264–9268 (1986).

32. Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol*. **31**, 827–832 (2013).
33. Nakagata, N., Okamoto, M., Ueda, O. & Suzuki, H. Positive effect of partial zona-pellucida dissection on the in vitro fertilizing capacity of cryopreserved C57BL/6J transgenic mouse spermatozoa of low motility. *Biol. Reprod*. **57**, 1050–1055 (1997).
34. Schmid-Burgk, J. L. et al. OutKnocker: a web tool for rapid and simple genotyping of designer nuclease edited cell lines. *Genome Res*. **24**, 1719–1723 (2014).
35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
36. Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J. L. CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS ONE* **10**, e0124633 (2015).
37. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
38. Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol*. **17**, 148 (2016).
39. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol*. **29**, 24–26 (2011).
40. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
41. Tran, N. H. et al. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat. Methods* **16**, 63–66 (2019).
42. Tran, N. H., Zhang, X., Xin, L., Shan, B. & Li, M. De novo peptide sequencing by deep learning. *Proc. Natl Acad. Sci. USA* **114**, 8247–8252 (2017).
43. Najm, F. J. et al. Rapid and robust generation of functional oligodendrocyte progenitor cells from epiblast stem cells. *Nat. Methods* **8**, 957–962 (2011).
44. Lager, A. M. et al. Rapid functional genetics of the oligodendrocyte lineage using pluripotent stem cells. *Nat. Commun*. **9**, 3708 (2018).
45. Hagemann, T. L. et al. Antisense suppression of glial fibrillary acidic protein as a treatment for Alexander disease. *Ann. Neurol*. **83**, 27–39 (2018).
46. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. **10**, R25 (2009).
47. Glascock, J. J. et al. Delivery of therapeutic agents through intracerebroventricular (ICV) and intravenous (IV) injection in mice. *J. Vis. Exp*. **56**, 2968 (2011).

Acknowledgements This research was supported, in part, by grants from the National Institutes of Health R01NS093357 (P.J.T.), T32GM007250 (M.S.E, Z.S.N. and K.C.A), F30HD084167 (Z.S.N.), F30HD096784 (K.C.A.) and T32NS077888 (K.C.A.); the New York Stem Cell Foundation (P.J.T); the European Leukodystrophy Association (P.J.T.); and philanthropic contributions from the Research Institute for Children's Health and the Geller, Goodman, Fakhouri, Long, Matreyek, Peterson and Weidenthal families. Additional support was provided by the Genomics, Small Molecule Drug Development, Transgenic and Rodent Behavioral core facilities of the Case Western Reserve University (CWRU) Comprehensive Cancer Center (P30CA043703), the Data Analytics Core of the Department of Population and Quantitative Health Sciences at CWRU, the CWRU Light Microscopy Imaging Center (S10OD016164), the electron microscopy division of the Cleveland Clinic Lerner Research Institute Imaging Core, and the University of Chicago Genomics Facility. We thank L. Landmesser, P. MacFarlane, R. Miller, P. Scacheri, T. Wynshaw-Boris, B. Clayton, S. Edelheit, A. Miron, H. Arakawa, P. Philippidou, A. Vagnozzi, L. Hu, E. Cohn, M. Scavuzzo, C. Allan and J. Cregg for technical assistance, discussion and review of the manuscript.

Author contributions M.S.E. and P.J.T. conceived and managed the overall study. H.E.S and M.S.E. maintained the animal colonies and tracked survival. M.S.E. captured video recordings. L.B. and M.S.E. designed and tested sgRNAs. S.H. performed data analysis for CRISPR off-target assessments. D.F.L., R.A.C. and W.J. performed zygote electroporation and oviduct transfers. H.E.S., B.S.N., K.C.A. and L.B. performed western blot experiments and protein quantification. D.M.S. and H.E.S. performed mass spectrometry sample preparation and analysis. B.E.P., L.B., B.S.N. and K.C.A. performed RT–qPCR. M.M., B.S.N, L.B., H.E.S., A.S.G. and M.S.E. generated and quantified the immunohistochemistry data. Y.M.-H. performed optic nerve electrophysiology studies and analysed the data. Y.M.-H., M.H., M.S.E. and H.E.S. processed samples for electron microscopy. Y.M.-H. analysed and quantified electron microscopy images. M.S.E., K.C.A., B.S.N. and L.B. performed rotarod and open-field experiments. M.S.E., B.S.N., K.C.A. and H.E.O. generated and characterized iPS cells and OPCs in vitro. H.T.Z. and A.S. generated *Hdac2*-targeting ASO data. B.E.P. and F.R. designed and characterized *Plp1*-targeting ASOs, tested tolerability in adult mice, recommended the use of ASOs, and contributed to the study design and interpretation of results in the ASO-treated disease model. M.S.E. performed ASO injections in *jimpy* mice. Y.M.-H. and B.S.N. performed all respiratory evaluations and analysed the data. Z.S.N. contributed key components to experimental design, data analysis and manuscript composition. M.S.E. and L.B. performed statistical analyses. M.S.E., M.M., Y.M.-H., L.B. and P.J.T. assembled figures. M.S.E. and P.J.T. wrote the manuscript with input from all authors.

Competing interests P.J.T. and M.S.E. are listed as inventors on pending patent claims (PCT/US2017/064870) filed by CWRU covering methods of PLP1 suppression. P.J.T. is a co-founder and consultant for Convelo Therapeutics, which has licensed patents unrelated to the current study from CWRU inventors (P.J.T., M.S.E., Z.S.N. and M.M.). P.J.T. and CWRU retain equity in Convelo Therapeutics. P.J.T. is a consultant and on the Scientific Advisory Board of Cell Line Genetics, which performed karyotyping in this study. P.J.T. is Chair of the Scientific Advisory Board (volunteer position) for the Pelizaeus-Merzbacher Disease Foundation. B.E.P., H.T.Z., A.S. and F.R. are employees of Ionis Pharmaceuticals. No other authors declare competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2494-3.

Correspondence and requests for materials should be addressed to P.J.T.

Peer review information *Nature* thanks Evan Goldstein and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Predicted target sites for sgRNA 3

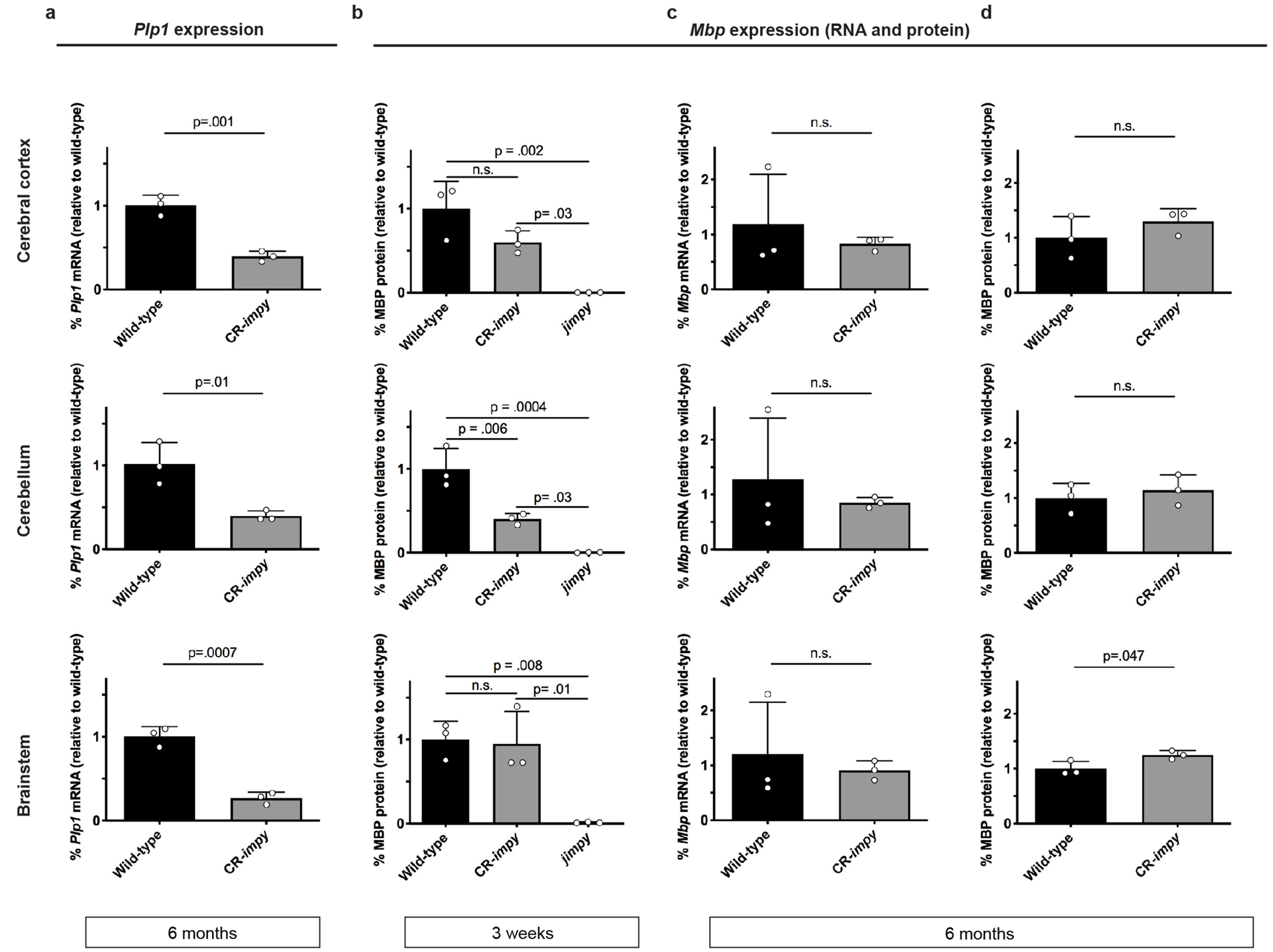
On-target site	Indel
chrX: 136832074-136832096	CR-*impy* deletion
Potential off-target sites	**Indel**
chr2: 122350283-122350305	-
chr13: 49148131-49148153	-
chr12: 36936047-36936069	-
chr6: 8903976-8903998	-
chr14: 89696741-89696763	-
chr7: 66891866-66891888	-
chr16: 92395171-92395193	-
chrX: 73879411-73879433	-
chr4: 101151689-101151711	-
chr19: 15977728-15977750	-
chr17: 85286736-85286758	-
chr1: 111378802-111378824	-
chr3: 64975280-64975302	-
chr15: 86959198-86959220	-
chr5: 141050321-141050343	-
chr15: 95488668-95488690	-
chr17: 74147915-74147937	-
chr14: 121445135-121445157	-
chr10: 77270008-77270030	-
chr19: 32474754-32474776	-
chr6: 53448359-53448381	-
chr5: 134254786-134254808	-
chr5: 34115009-34115031	-
chr11: 76791904-76791926	-
chr2: 41404945-41404967	-
chr9: 120474214-120474236	-
chr2: 166560989-166561011	-
chr14: 117904354-117904376	-
chr15: 87946536-87946558	-
chr3: 60096145-60096167	-
chr1: 189602230-189602252	-
chr10: 88264444-88264466	-
chr12: 95242735-95242757	-
chr9: 106347977-106347999	-
chr15: 91779810-91779832	-
chr9: 25058414-25058436	-
chr4: 46593785-46593807	-
chr7: 118420165-118420187	-
chr9: 96967183-96967205	-
chr2: 155735131-155735153	-
chr16: 64752130-64752152	-
chr13: 38607903-38607925	-
chr5: 111154824-111154846	-
chr19: 36554294-36554316	-
chr1: 45740662-45740684	-
chr3: 78025787-78025809	-
chr6: 30319176-30319198	-
chr9: 21496763-21496785	-
chr6: 90225254-90225276	-

Predicted target sites for sgRNA 7

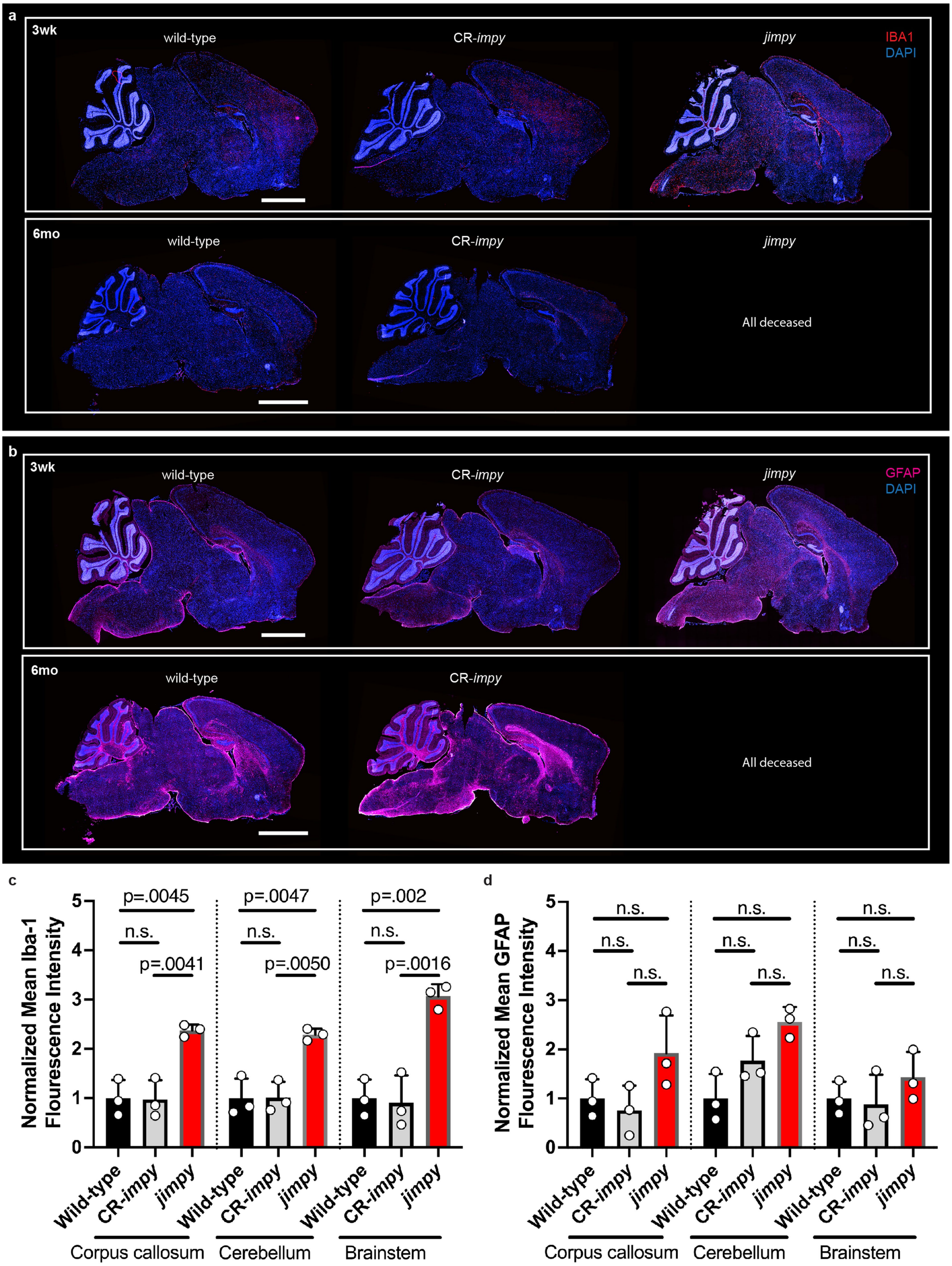
On-target site	Indel
chrX: 136832001-136832023	CR-*impy* deletion
Potential off-target sites	**Indel**
chr6: 124587509-124587531	-
chr17: 81109628-81109650	-
chr5: 31245537-31245559	-
chr7: 117170171-117170193	-
chr7: 27808065-27808087	-
chr5: 123615487-123615509	-
chr4: 70289082-70289104	-
chr12: 66380418-66380440	-
chr11: 40896483-40896505	-
chr17: 32871777-32871799	-
chr18: 81751517-81751539	-
chr12: 13393396-13393418	-
chr1: 173579543-173579565	-
chr17: 55952630-55952652	-
chr7: 73141599-73141621	-
chr13: 70096648-70096670	-
chr1: 191510521-191510543	-
chr1: 82207126-82207148	-
chr3: 87376275-87376297	-
chr17: 66699234-66699256	-
chr2: 131423764-131423786	-
chr3: 152532012-152532034	-
chr13: 29505709-29505731	-
chr17: 20996364-20996386	-
chr6: 129031384-129031406	-
chr6: 136770085-136770107	-
chr4: 96900332-96900354	-
chr6: 7985354-7985376	1bp insertion
chr4: 33744227-33744249	-
chr4: 138806927-138806949	-
chr19: 52530396-52530418	-
chr8: 120455576-120455598	-
chr2: 124374333-124374355	-
chr10: 77242641-77242663	-
chr2: 29431651-29431673	-
chr6: 77211332-77211354	-
chr1: 133219251-133219273	-
chr1: 168417730-168417752	-
chr11: 95154989-95155011	-
chr7: 131374279-131374301	-
chr9: 106450035-106450057	-
chr7: 139001357-139001379	-
chr11: 99410761-99410783	1bp insertion
chr17: 31067019-31067041	-
chr12: 100612177-100612199	-
chr9: 79131521-79131543	-
chr2: 10352651-10352673	-
chr14: 105403488-105403510	-
chr18: 14043380-14043402	-

Extended Data Fig. 1 | See next page for caption.

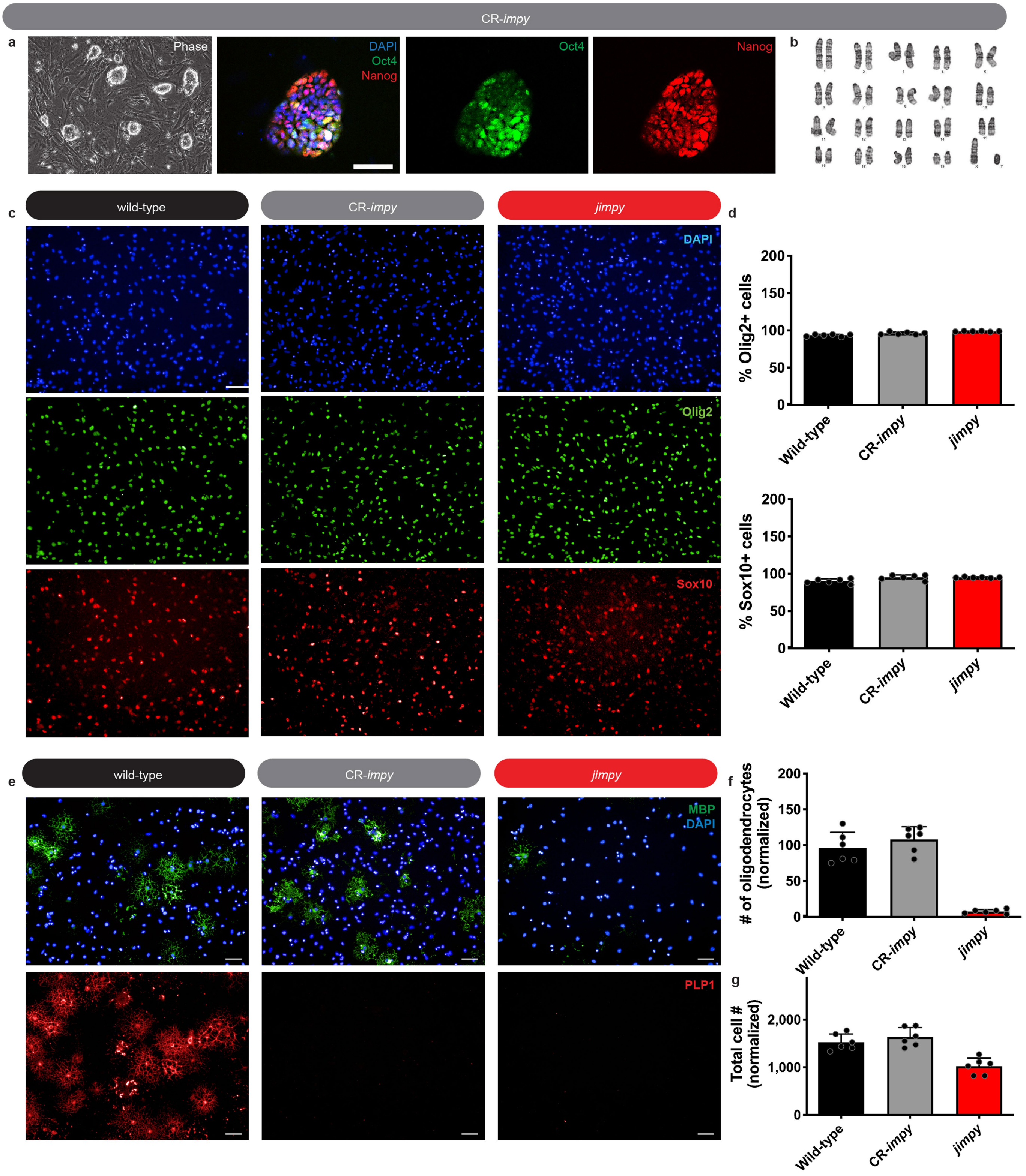
Extended Data Fig. 1 | CRISPR nuclease induction of *Plp1* frameshift mutations in *jimpy* with high accuracy. a, Annotated Sanger sequencing traces of wild-type, *jimpy*, and CR-*impy* mice showing the complex, frameshift in *Plp1* exon 3 from dual cutting of CRISPR/spCas9 sgRNAs in CR-*impy* mice as well as the *jimpy* point mutation in intron 4. sgRNA 3 and 7 sequences outlined by black boxes with the predicted double strand break site shown a black arrow. **b**, Table showing the top predicted on- and off-target sites for sgRNAs 3 and 7. CRISPR-induced indels were detected by whole genome sequencing of the CR-*impy* founder and three independent CR-*impy* F2 generation males, and consisted of an on-target 80bp complex deletion (CR-*impy* deletion) in exon 3 of *Plp1* (green), an off-target 1 bp insertion in chromosome 6 (red), and an off-target 1 bp insertion in chromosome 11 (yellow). **c**–**e**, Integrative Genomics Viewer browser images showing aligned reads for the CR-*impy* founder, the *jimpy* control, and three CR-*impy* F2 males along with the detected indels at the on-target locus at exon 3 of *Plp1* on chromosome X (**c**), and off-targets on chromosome 6 (**d**) and chromosome 11 (**e**) depicted by the dashed green, red, and yellow boxes, respectively. sgRNA 3 or sgRNA 7 targeted sequences are depicted by black bars.



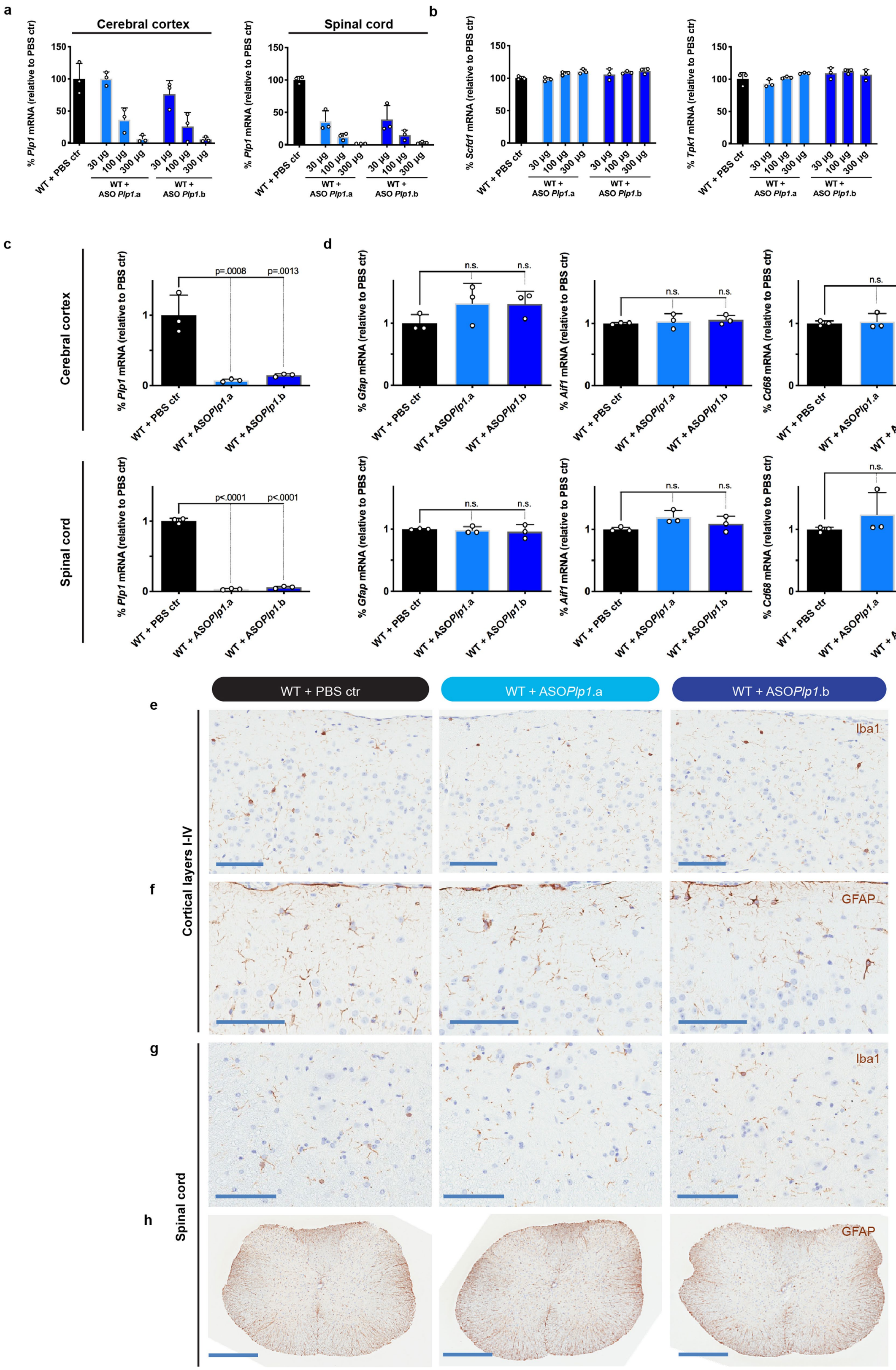
Extended Data Fig. 2 | CRISPR-mediated suppression of *Plp1* in *jimpy* mice increases *Mbp* expression across multiple CNS regions. a, RT–qPCR data showing the levels of *Plp1* transcript at 6 months ($n = 3$ mice). **b**, western blot data demonstrating the levels of MBP protein at 3 weeks ($n = 3$ mice). **c**, RT–qPCR data showing the levels of *Mbp* transcript at 6 months ($n = 3$ mice). **d**, western blot data demonstrating the levels of MBP protein at 6 months ($n = 3$ mice). Individual data points represent the mean value of 4 technical replicates for each biological replicate (**a**, **c**) or independent biological replicates (**b**, **d**). Biological replicates (individual mice) indicated by open circles. Graph bars indicate mean ± standard deviation. p-values calculated using one-way ANOVA with Tukey correction at 3 weeks or two-way, an unpaired two-sided *t*-test at later time points. p-values stated for $P < 0.1$, otherwise not significant (n.s). See Supplementary Data 2 for full western blot images for all samples.



Extended Data Fig. 3 | CRISPR-mediated suppression of *Plp1* in *jimpy* mice reduces markers of activated microglia and astrocytes. a, Immunohistochemical images of whole-brain sagittal sections showing Iba1$^+$ microglia (red) and DAPI$^+$ nuclei (blue) across genotypes. Scale bar, 2mm. **b**, Immunohistochemical images of whole-brain sagittal sections showing GFAP$^+$ astrocytes (red) and DAPI$^+$ nuclei (blue) staining across genotypes. Scale bar, 2mm. **c**, **d**, Normalized mean signal intensity of (**c**) Iba1$^+$ microglia and (**d**) GFAP$^+$ astrocytes across genotypes and CNS regions ($n = 3$ mice). Biological replicates (individual mice) indicated by open circles. Graph bars indicate mean ± standard deviation. p-values calculated using one-way ANOVA with Tukey correction. p-values stated for $P < 0.1$, otherwise not significant (n.s). See Supplementary Data 3-5 for representative source images of Iba-1 and GFAP staining.

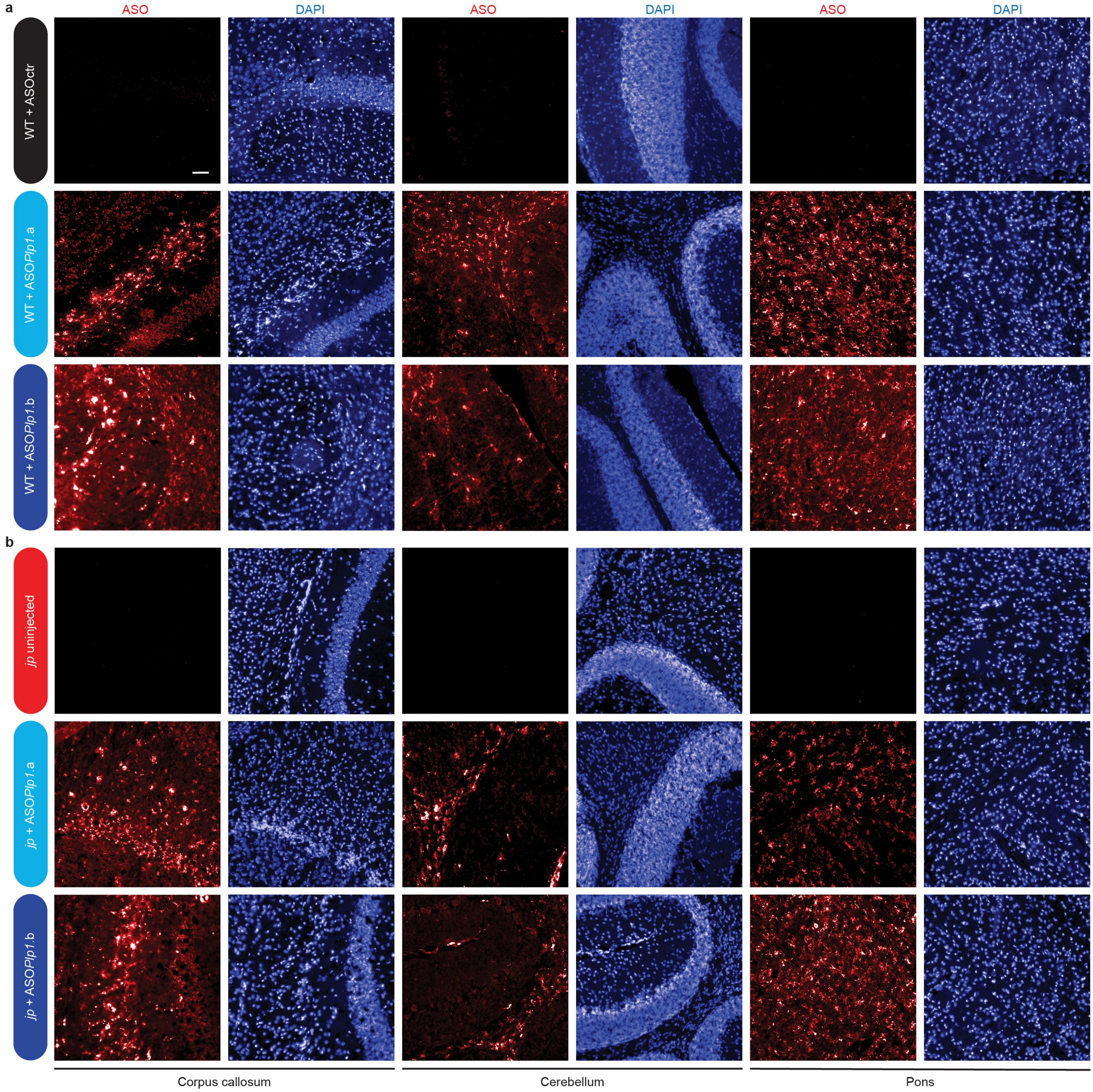


Extended Data Fig. 4 | *Plp1* suppression in *jimpy* OPCs rescues survival of differentiating oligodendrocytes in vitro. **a**, Phase and immunocytochemistry images of Oct4$^+$ and Nanog$^+$ iPS cells, along with DAPI$^+$ nuclei and **b**, normal karyotype of a CR-*impy* iPS cell line used to generate OPCs. Scale bar, 50µm. **c**, Immunocytochemistry images showing Olig2$^+$ and Sox10$^+$ cells in OPC cultures, along with DAPI$^+$ nuclei, derived from iPS cells. Scale bar, 100µm. **d**, Percentage of Sox10$^+$ and Olig2$^+$ cells in OPC cultures. **e**, Immunocytochemistry images of MBP$^+$ and PLP$^+$ oligodendrocytes. **f**, **g**, Quantification of (**f**) MBP$^+$ oligodendrocytes and (**g**) total cell number (DAPI$^+$ nuclei) from iPS-cell-derived OPCs differentiated in vitro for 3 days. Scale bar, 50µm. Technical replicates (individual wells) for a single cell line per genotype indicated by black circles. Graph bars indicate mean ± standard deviation.

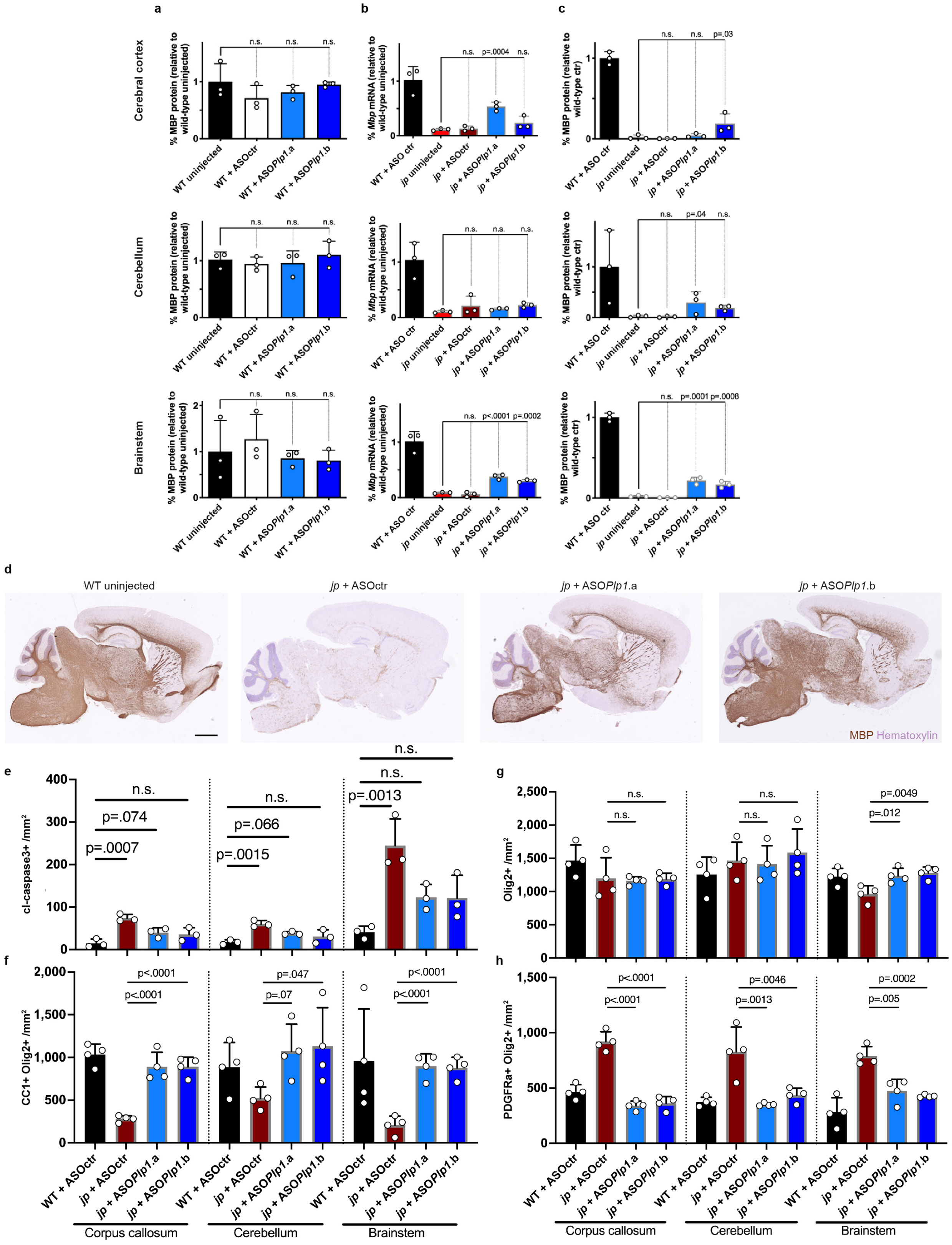


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | *Plp1*-targeted ASOs do not suppress off-target transcripts or activate glial cells. **a**, **b**, RT–qPCR data showing the level of (**a**) *Plp1* transcript levels or (**b**) expression levels of off-target transcripts (up to 3 base mismatches) in the spinal cord for *Plp1*-tageting ASOs, including *Xylt1* (off-target for ASO *Plp1*.a), *Scfd1*, or *Tpk1* (off-targets for ASO *Plp1*.b), 2 weeks post-injection of *Plp1*-targeting ASOs (30µg, 100µg, and 300µg doses) or PBS control in 8 week old adult wild-type (wt) mice ($n = 3$ mice). **c**, **d**, RT–qPCR data showing *Plp1* transcript levels or tolerability by expression levels of *Gfap, Aif1*, and *Cd68* transcripts in the cerebral cortex and spinal cord, 8 weeks post-injection with the indicated ASOs (300µg dose) or PBS control in 8 week old wild-type mice ($n = 3$ mice). **e**–**h**, Immunohistochemistry images with haematoxylin counterstain showing Iba1$^+$ or GFAP$^+$ astrocytes in **e**, Cortical layers I-IV (Iba1), (**f**) cortical layers I-III (GFAP), (**g**) spinal cord dorsal horn grey/white matter intersection (Iba1), and (**h**) spinal cord (GFAP), 8 weeks post-injection with the indicated ASOs (300µg dose) or PBS control in 8 week old wild-type mice. Scale bar, 500µm. Biological replicates (individual mice) indicated by open circles, representing the mean value of 3 technical replicates. Graph bars indicate mean ± standard deviation. p-values calculated using one-way ANOVA with Dunnett's correction for multiple comparisons. p-values stated for $P < 0.1$, otherwise not significant (n.s).

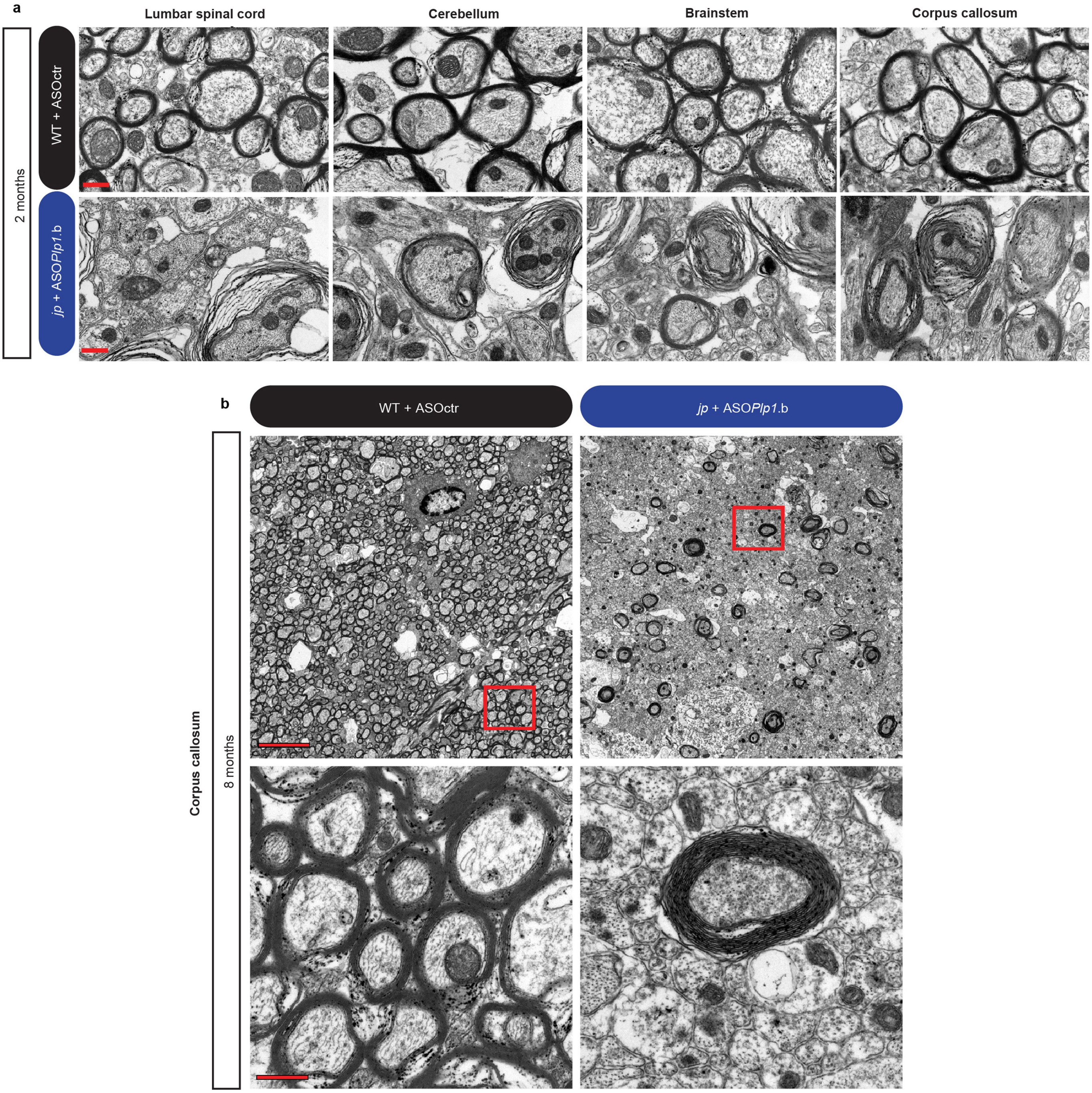


Extended Data Fig. 6 | *Plp1*-targeted ASOs distribute widely throughout the CNS after ICV injection in postnatal mice. a, **b**, Immunohistochemical images of brain sagittal sections showing ASO$^+$ staining and DAPI$^+$ nuclei (blue) of WT + ASO*Plp1.a*, WT + ASO*Plp1.b* and WT uninjected (**a**) or jp + ASO*Plp1.a*, jp + ASO*Plp1.b* and *jimpy* uninjected mice (**b**), 3 weeks post-ASO injection (30 µg dose at birth). Scale bar, 50 µm.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | *Plp1*-targeting ASOs increase *Mbp* expression and rescue oligodendrocyte numbers in *jimpy* mice. a, Western blot data showing the level of MBP protein ($n = 3$ mice). **b**, RT–qPCR data showing the level of *Mbp* transcript ($n = 3$ mice). **c**, Western blot data showing the level of MBP ($n = 3$ mice). **d**, Immunohistochemistry images with haematoxylin counterstain of whole brain sagittal sections showing MBP$^+$ myelin. Scale bar, 1mm. **e**, Quantification of cleaved-caspase 3$^+$ apoptotic cells ($n = 3$ mice). **f**, Quantification of CC1$^+$/Olig2$^+$ oligodendrocytes ($n = 4$ mice). **g**, Quantification of the number of Olig2$^+$ glial lineage cells ($n = 4$ mice). **h**, Quantification of the number of PDGFRα$^+$/Olig2$^+$ OPCs ($n = 4$ mice). All data collected at 3 weeks post-ASO injection (30μg dose at birth). Individual data points represent the mean value of 4 technical replicates for each biological replicate (individual mice) (**b**) or independent biological replicates (individual mice) (**a**, **c**–**h**), indicated by open circles. Graph bars indicate mean ± standard deviation. p-values calculated using one-way ANOVA with Dunnett's correction for multiple comparisons. p-values stated for $P < 0.1$, otherwise not significant (n.s). See Supplementary Data 4 for full western blot images for all samples.



Extended Data Fig. 8 | *Plp1*-targeted ASOs induce sustained myelination throughout the neuraxis in *jimpy* mice. a, b, Electron micrograph images showing myelination of WT + ASOctr or *jp* + ASO*Plp1.b* at 2 months (**a**) and 8 months (**b**). For **a**, scale bar, 0.5 µm. In **b**, the bottom panel is a higher magnification of red boxed area in the top panel. Scale bars, 5 µm (top) and 0.5 µm (bottom).

nature research

Corresponding author(s): Paul J. Tesar

Last updated by author(s): Jun 6, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Images were acquired with Leica Application Suite X, Hamamatsu NDP 2.0, or Perkin Elmer Operetta Harmony software. Videos were acquired using an Apple iPhone. Optic nerve conduction velocity was recorded using AxoScope software(Molecular Devices). Behavioral measurements were recorded using ANY-maze software version 5.0 (open field) and Rota Rod Rotomax 5 (rotarod). Breathing was recorded on the IOX2 software (Emka).
Data analysis	Graphpad Prism was used to generate graphs and perform statistics. Adobe Photoshop, NIH ImageJ, and Perkin Elmer Harmony and Columbus software were used for calculations and cell counting. spCas9 CRISPR sgRNA design tool at crispr.mit.edu was used to design sgRNAs. CRISPR-induced indels were analyzed using the OutKnocker tool at outknocker.org, GATK RealignerTargetCreator, IndelRealigner (version 3.3-2-gec30cee), Blat (v. 36x2), CCTop, RGEN Cas-OFFinder, CRISPOR, and the Integrative Genomics Viewer. Bowtie aligner 58 was used to identify putative ASO off-target transcript sequences. Adobe Photoshop and Illustrator were used to assemble images. Blots were analyzed with the Odyssey Fc imaging system (Li-Cor). LC-MS/MS data was analyzed using BioinformaticsSolutions PeaksStudio software.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated or analyzed during this study are included in this article and its supplementary information files. Source data for animal survival cohorts in Figs. 1b, k-l, and 3b, 4a-b are provided in Supplementary Data 1 and 6. Raw annotated western blot images for Extended Data Fig. 2b, d and Extended Data Fig. 7a, c are provided as Supplementary Data 2 and 7. Source data for all graphs are provided as separate Excel files. Animals and iPSC lines are available from P.J.T. upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical test was used to predetermine sample size. Instead, sample sizes were rationalized by considering sufficient replication (weighing the level of biological variation) as well as censoring (due to tissue harvesting at pre-determined time-points and inadvertent losses).
Data exclusions	All data points were included in analyses except for certain animals that were censored from survival analyses to use in pre-determined terminal assays. Metadata for all mice in this study including censoring of animals in the survival analyses are found in Supplementary Figs. 1 and 3.
Replication	The ASO therapeutic response was tested with two independent ASOs and all data were replicated.
Randomization	Sample allocation was not random. Instead, biological controls were employed in all experiments.
Blinding	Investigators were blinded to animal genotype at the time of ASO injection. Investigators were blinded to genotype and treatment for immunohistochemistry quantifications. For other experiments (i.e. animal behavior, electrophysiology, and respiratory analysis) blinding was not possible due to the overt jimpy phenotype.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☐	☒	Eukaryotic cell lines
☒	☐	Palaeontology
☐	☒	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used

Primary antibodies used for IHC: mouse anti-MBP (2μg/mL; 808401, Biolegend; RRID:AB_2564741), rabbit anti-MBP (1:1000; Abcam, ab40390; RRID:AB_1141521), rabbit anti-MyRF polyclonal antibody (1:500; kindly provided by Dr. Michael Wegner), goat anti-SOX10 (0.4μg/mL; AF2864, R&D Systems; RRID:AB_442208), rabbit anti-GFAP (1:1000; Z0334, Dako; RRID:AB_10013382), goat anti-IBA1 (0.1mg/mL; ab5076, Abcam), rabbit anti-IBA1 (1:2000; 019–19741, WAKO; RRID:AB_839504), rabbit anti-ASO (1:2500; Ionis Pharmaceuticals, Carlsbad, CA), rabbit anti-HDAC2 (1:250; Abcam, ab16032; RRID:AB_2118543), mouse anti-APC/CC1 (2.5 μg/ml; ab16794, Abcam; RRID:AB_443473), mouse anti-APC/CC1 (1:250; MABC200, Millipore; RRID:AB_11203645), rat anti-NG2 (25 μg/mL; MAB6689, R&D Systems; RRID:AB_10890940), goat anti-PDGFRα (1:500; AF1062, R&D systems; RRID:AB_2236897), and rabbit anti-OLIG2 (1:250; 13999-1-AP, ProteinTech; RRID:AB_2157541).

Primary antibodies used for western blot: mouse anti-MBP antibody (1μg/mL; 808401, Biolegend; RRID:AB_2564741) and rat anti-PLP antibody (1:1000; clone AA3, Lerner Research Institute Hybridoma Core, Cleveland, OH).

Primary antibodies used for ICC: mouse anti-MBP (1:500; 808401, Biolegend; RRID:AB_2564741), rat anti-PLP (1:5000; clone AA3, Lerner Research Institute Hybridoma Core, Cleveland, OH), goat anti-SOX10 (2μg/mL; AF2864, R&D Systems; RRID:AB_442208), rabbit anti-OLIG2 (1:1000; 13999-1-AP, ProteinTech; RRID:AB_2157541), rabbit anti-NANOG (0.4μg/mL; AB21624, Abcam; RRID:AB_446437), mouse anti-OCT3/4 (0.4μg/mL; SC-5279, Santa Cruz; RRID:AB_628051).

Validation	Primary antibodies used in this study are well accepted in the field and purchased from reputable suppliers with provided quality control metrics.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Mouse iPSC lines were generated in-house
Authentication	Cells lines were genotyped, karotyped, and stained for canonical markers of OPCs and iPSCs.
Mycoplasma contamination	Laboratory cell lines are routinely tested for mycoplasma contamination with consistently negative results.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Male jimpy mice (B6CBACa-Aw-J/A-Plp1jp EdaTa/J; RRID:IMSR_JAX:000287), CRISPR modified jimpy (CR-impy) mice (this paper) and wild-type controls. All mice were on a B6CBACa background.
Wild animals	No wild animals were used in this study.
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	All procedures were in accordance with the National Institutes of Health Guidelines for the Care and Use of Laboratory Animals and were approved by the Case Western Reserve University Institutional Animal Care and Use Committee (IACUC).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Keratins are asymmetrically inherited fate determinants in the mammalian embryo

https://doi.org/10.1038/s41586-020-2647-4

Received: 23 October 2019

Accepted: 30 July 2020

Hui Yi Grace Lim[1], Yanina D. Alvarez[1], Maxime Gasnier[1], Yiming Wang[2], Piotr Tetlak[1], Stephanie Bissiere[1], Hongmei Wang[2], Maté Biro[3] & Nicolas Plachta[1,4]✉

To implant in the uterus, the mammalian embryo first specifies two cell lineages: the pluripotent inner cell mass that forms the fetus, and the outer trophectoderm layer that forms the placenta[1]. In many organisms, asymmetrically inherited fate determinants drive lineage specification[2], but this is not thought to be the case during early mammalian development. Here we show that intermediate filaments assembled by keratins function as asymmetrically inherited fate determinants in the mammalian embryo. Unlike F-actin or microtubules, keratins are the first major components of the cytoskeleton that display prominent cell-to-cell variability, triggered by heterogeneities in the BAF chromatin-remodelling complex. Live-embryo imaging shows that keratins become asymmetrically inherited by outer daughter cells during cell division, where they stabilize the cortex to promote apical polarization and YAP-dependent expression of CDX2, thereby specifying the first trophectoderm cells of the embryo. Together, our data reveal a mechanism by which cell-to-cell heterogeneities that appear before the segregation of the trophectoderm and the inner cell mass influence lineage fate, via differential keratin regulation, and identify an early function for intermediate filaments in development.

The development of multicellular organisms requires the specification of diverse lineages from a small group of cells within the embryo. During mammalian development, the first lineage segregation produces the pluripotent inner cell mass (ICM), which forms the fetus and primitive endoderm, and the outer trophectoderm that forms the placenta[1]. How these lineages are specified remains unclear. The 'inside–outside' model suggests that lineage fates are specified by local signals after cells segregate into inner-outer positions[3]. Contrary to this model, heterogeneities in histone modifications[4], transcription factor dynamics[5], non-coding RNA localization[6], and gene expression[7–10] appearing as early as the four-cell stage bias the acquisition of pluripotent and trophectoderm fates, yet the mechanism is unclear.

By contrast, the 'cell polarity' model proposes that asymmetric inheritance of polarity components during cell division specifies distinct fates[11]. Some suggested that this relies on the asymmetric inheritance of the apical domain, which forms at the eight-cell stage before divisions segregating inner and outer cells, via enrichment of F-actin, PAR6 and aPKC at the apical cortex[12–14]. This would be consistent with studies showing that apical polarity at later stages promotes nuclear retention of the transcription factor YAP, which supports high expression of CDX2, a key transcription factor that specifies trophectoderm identity[15]. However, live-embryo imaging revealed that the apical domain disassembles from the cortex before division, instead of being directly inherited[16]. Therefore, it remains unclear whether other polarized components function as asymmetrically inherited fate determinants during mammalian development, similar to those in non-mammalian embryos[2], and how they relate to heterogeneities at earlier stages.

The cytoskeleton is not only composed of F-actin and microtubules, but also of various intermediate filaments[17]. During preimplantation development, keratins are the only cytoplasmic intermediate filaments that are expressed[18–20]. Keratins regulate polarity, signalling and mechanics in epithelial tissues[17], and have traditionally served as markers of trophectoderm[19]. Moreover, keratin knockouts display trophoblast fragility, placental bleeding and lethality after implantation[21–23]. Yet, keratin functions during preimplantation development remain unknown.

To study their functions, we performed immunofluorescence for keratins 8 and 18 (K8 and K18), the subtypes that are predominantly expressed during preimplantation stages[24]. In contrast to F-actin and microtubules, keratins are the first cytoskeletal component that displays cell-to-cell variability during development[25,26] (Fig. 1a). Although keratins are well-established markers of trophectoderm, the first filaments are already detected in a subset of cells of the eight-cell mouse embryo, before lineage segregation, with a similar pattern in the human embryo (Fig. 1, Extended Data Fig. 1a–c). The proportion of cells assembling filaments increases over time (Fig. 1b–d), and by the blastocyst stage the trophectoderm is covered by a dense network, whereas the ICM is devoid of keratins[18,27] (Fig. 1b–c, Extended Data Fig. 1d, Supplementary Video 1). Thus, variability in the assembly of keratin filaments establishes differences in cytoskeletal organization before segregation of the ICM and trophectoderm.

---

[1]Institute of Molecular and Cell Biology, ASTAR, Singapore, Singapore. [2]State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Science, Beijing, China. [3]EMBL Australia, Single Molecule Science Node, School of Medical Sciences, University of New South Wales, Sydney, New South Wales, Australia. [4]Department of Cell and Developmental Biology and Institute for Regenerative Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ✉e-mail: plachtalab@gmail.com

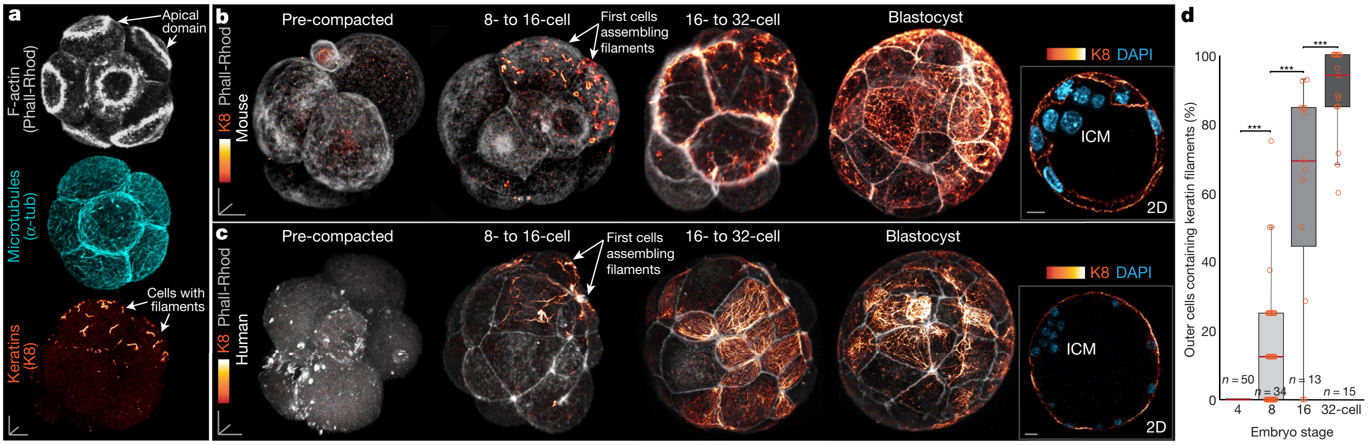


Fig. 1 | Keratin filaments display cell-to-cell variability before lineage segregation in the mouse and human embryo. a, All cells of the 8-cell embryo show similar F-actin and microtubule organization, yet only a subset assembles keratin filaments. Data are from five independent experiments. **b**, **c**, Mouse and human embryos at several developmental stages. Keratin filaments initially assemble in a subset of cells at the 8- to 16-cell stage, before inner–outer cell segregation. Blastocysts show dense keratin networks in the trophectoderm, but not in the ICM. Insets show middle 2D views. **d**, Quantification of keratin filament-forming outer cells of the embryo throughout preimplantation development. In box plots, the centre line is the median, box edges show upper and lower quartiles and whiskers represent the range. $^{***}P < 0.0001$, analysis of variance (ANOVA) test. Data are from three independent experiments. Scale bars, 10 μm.

We next microinjected embryos with mRNA for fluorescently labelled K18 (K18-Emerald), which display expression patterns that resemble endogenous keratins (Extended Data Fig. 1e–g). Live-imaging and immunofluorescence show that keratin filaments start to assemble in the sub-cortical and cortical regions during interphase, before apical domain formation (Extended Data Fig. 2a, b, Supplementary Video 2). The size of keratin filaments increases over time, and their motion is unconfined with an average speed of 0.45 ± 0.08 μm min^{-1} (Extended Data Fig. 2c–f), similar to measurements in cultured cells[28]. However, when the apical domain forms, keratins become more static and enriched at this structure, which suggests that keratins anchor to the apical domain (Fig. 2a, Extended Data Fig. 2a–b, f–g, Supplementary Video 3). Consistently, treatment with cytochalasin D blocks the formation of apical domains and shifts keratin localization to more uniform along the cortex and cytoplasm (Fig. 2a, Extended Data Fig. 2h). By contrast, acute treatment with SiR-Actin, which stabilizes F-actin, increases the density of F-actin at the apical domain and keratin apical polarization (Fig. 2a, Extended Data Fig. 2h). Therefore, the apical domain serves as scaffold to enrich keratins apically during interphase.

In many tissues, keratins anchor to the cortex via desmosomes[17]. Although mature desmosomes assemble by blastocyst stage, desmosome components are expressed in eight-cell embryos[29,30]. Immunofluorescence for the endogenous desmosome components plakoglobin, plakophilin and desmoglein2 reveals their localization to the apical domain (Extended Data Fig. 3a–c). Imaging fluorescently labelled desmoglein2 and K18 in live embryos confirms this pattern (Extended Data Fig. 3d–f). Furthermore, downregulation of desmosome proteins reduces keratin apical polarization (Extended Data Fig. 3g, h). Thus, desmosome components link keratin filaments to the apical domain.

When cells enter mitosis, they largely disassemble their apical domain[16], cortical microtubules, and desmosome components (Extended Data Figs. 3e, 4a). By contrast, keratins are stably retained within mitotic cells across different developmental stages, consistent with fluorescence recovery after photobleaching (FRAP) revealing a larger immobile fraction for keratins than actin (Extended Data Fig. 4b–d). Notably, live imaging of embryos expressing K18-Emerald shows that keratin filaments become asymmetrically inherited by the outer daughter cell during divisions producing inner–outer daughters, and symmetrically inherited during divisions producing outer–outer cells (Fig. 2b, Extended Data Fig. 4e, f, Supplementary Video 4). We confirmed these inheritance patterns in non-injected embryos (Extended Data Fig. 5a). These findings establish keratins as an asymmetrically inherited component during cell division.

To explore the inheritance mechanism, we used short interfering RNAs (siRNAs) that target PARD6B, which prevent the formation of the apical domain without interfering with the completion of mitosis[16,31] (Extended Data Fig. 5b). Knockdown of PARD6B reduces keratin apical polarization in interphase (Fig. 2a) and causes a more symmetric inheritance, even during divisions producing inner–outer daughter cells (Fig. 2c), which indicates that apical polarization of keratins before division is required for their asymmetric inheritance. Tracking keratins throughout mitosis shows that they still retain a high apical polarization, even after apical domain disassembly (Extended Data Fig. 2a, f), which was confirmed via immunofluorescence and in human embryos (Extended Data Fig. 5a, c). This apical retention suggests that some property of the cell hinders keratin movement, as long polymers diffuse more slowly through dense entangled meshworks[32]. Consistently, we found that mitotic cells display a dense cytoplasmic F-actin meshwork similar to earlier stages[33], through which keratin filaments move (at 0.4 μm min^{-1}) (Fig. 2d, Extended Data Fig. 5d, e). The speed of keratin filaments is inversely proportional to their volume (Extended Data Fig. 5f), in line with polymer studies[34]. Disrupting the F-actin meshwork using cytochalasin D specifically during mitosis causes keratins to move faster and lose their apical localization (Extended Data Fig. 5g). Moreover, when cells are arrested in metaphase using MG132, keratins have a longer time to move through the F-actin meshwork and eventually lose their apical localization (Extended Data Fig. 5g). As the distance between the apical cortex and cytokinetic furrow is 23.5 ± 1.52 μm (mean ± s.d.), and the time between apical domain disassembly and cytokinesis is 34.9 ± 6.2 min (Extended Data Fig. 5h), our results indicate that the slow movement of keratins through the dense F-actin meshwork during the relatively short duration of mitosis biases their apical retention. Hence, we propose a mechanism for keratin inheritance in which (1) the apical domain provides a scaffold promoting apical localization of keratins during interphase via desmosome proteins, and (2) after disassembly of this scaffold in mitosis, the cytoplasmic meshwork of F-actin hinders keratin movement, maintaining most filaments apically and biasing their inheritance by the outer cell.

Given their asymmetric inheritance by outer cells, we explored whether keratins influence trophectoderm specification. Analysis of

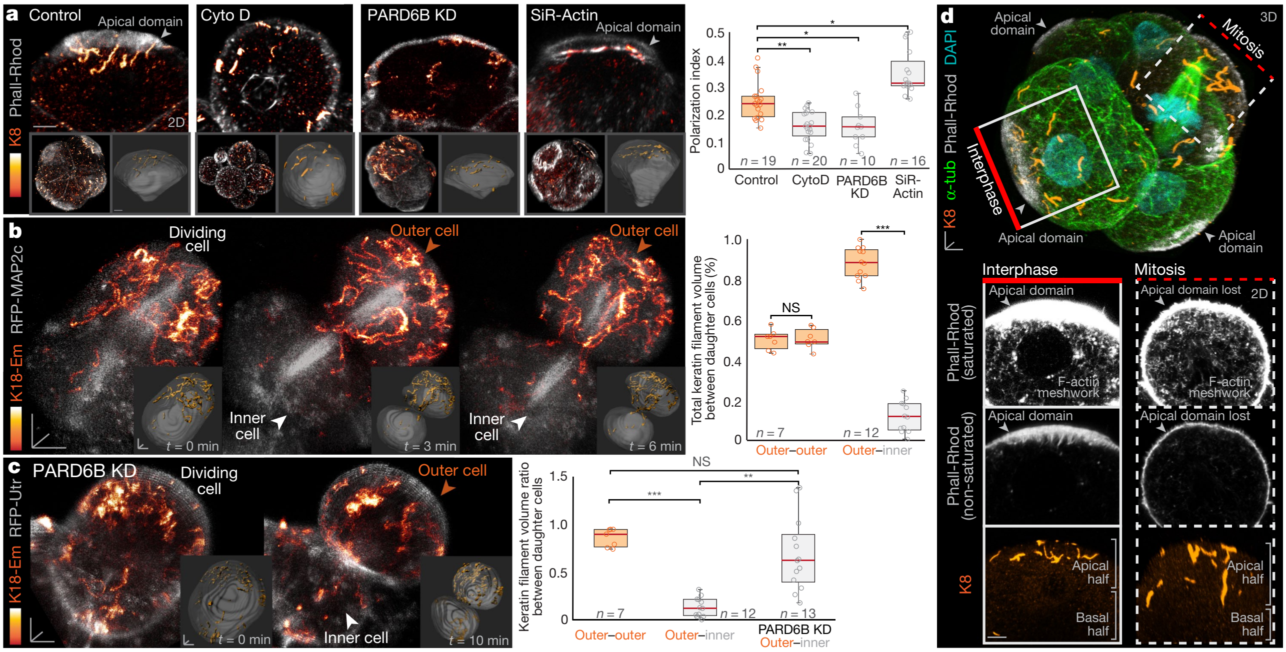


Fig. 2 | Keratin filaments are asymmetrically inherited during cell division. **a**, Keratin filaments are apically localized near the apical domain. Treatment with cytochalasin D and PARD6B knockdown (KD) reduce apical localization of keratins, whereas SiR-Actin increases it. Top panels show individual cells. Bottom panels show whole embryo (left) and computationally rendered filaments within each cell (right). Phall-Rhod, phalloidin-rhodamine. **$P = 0.004$; *$P = 0.02$ for PARD6B KD; *$P = 0.03$ for SiR-Actin; Kruskal–Wallis test. **b**, Imaging fluorescently tagged keratins within the live embryo reveals their asymmetric inheritance by the outer daughter during divisions producing inner and outer cells. Quantification shows inheritance patterns. RFP–MAP2c, red fluorescent protein (RFP)-tagged MAP2c. NS, not significant. ***$P < 0.0001$, Student's t-test. **c**, PARD6B knockdown shifts keratin inheritance from asymmetric to more symmetric in outer–inner divisions. ***$P = 0.0001$; **$P = 0.0009$; Kruskal–Wallis test. **d**, Immunofluorescence of eight-cell embryo highlights cytoskeletal organization in interphase and mitosis. While the apical domain and cortical microtubules become largely reorganized during mitosis, keratins retain their apical localization. 2D panels show loss of apical domain, but retention of a dense cytoplasmic F-actin meshwork and apically-localized keratins during mitosis. Data are from five independent experiments. In box plots, the centre line is the median, box edges show upper and lower quartiles and whiskers represent the range. Scale bars, 5 µm.

8- to 16-cell embryos shows that after division, cells inheriting keratins rapidly establish a dense network under the cortex (Fig. 3a). Although most 16-cell outer blastomeres reform an apical F-actin ring after division[16], only those that inherit keratins display higher levels of the apical polarity proteins PARD6B and PKCζ (Fig. 3b), and a larger immobile fraction of mRuby2-actin at this ring, compared to keratin-negative cells (Extended Data Fig. 6a–c). Furthermore, manipulation of actin stability per se using cytochalasin D reduces polarization, whereas stabilization with SiR-Actin increases polarization (Extended Data Fig. 6d). Knockdown of desmosome components also reduces actin stability (Extended Data Fig. 6a–c) and disrupts polarity (Extended Data Fig. 6e). Thus, keratins promote apical polarization by regulating the stability of F-actin.

Apical polarization is thought to oppose cell internalization and trigger YAP-dependent expression of CDX2 to establish trophectoderm identity[12,13]. Consistently, keratin-inheriting cells remain restricted to the outer layer, whereas most keratin-negative cells can undergo apical constriction[35] to form the ICM (Extended Data Fig. 6f, g, Supplementary Video 5). Keratin-inheriting cells also display the highest levels of nuclear YAP and CDX2, and the lowest levels of NANOG (Fig. 3b, c). Consistently, apical AMOT, which links apical polarity to YAP localization[36,37], is also enriched in these cells (Extended Data Fig. 6h). To test the role of keratins in trophectoderm specification, we combined siRNAs for K8 and K18, an approach that minimizes compensatory effects from weakly expressed keratins[38] and extensively eliminates the keratin network (Extended Data Fig. 7a, b). In contrast to cells inheriting keratins, K8/K18-knockdown cells display lower levels of apical PARD6B, PKCζ and AMOT, reduced nuclear expression of YAP and CDX2, and higher expression of NANOG (Fig. 3b, c, Extended Data Fig. 6h). CDX2 and NANOG levels in these knockdown cells are similar to inner cells, suggesting that K8/K18-knockdown cells are not yet specified to the trophectoderm (Fig. 3b, c). By contrast, co-injecting a high concentration of *K8* (also known as *Krt8*) and *K18* (*Krt18*) mRNA triggers the formation of a premature keratin network across all cells (Extended Data Fig. 7c), accompanied by widespread increase in CDX2 expression (Fig. 3d). Although the inner cells of these embryos inherit some overexpressed keratins, they still display low nuclear YAP and CDX2 levels, consistent with their lack of apical polarity (Extended Data Fig. 7d, e). Furthermore, keratin-positive cells in YAP-knockdown embryos are unable to maintain CDX2 expression, confirming that keratins regulate CDX2 via YAP (Extended Data Fig. 7f). Finally, microinjecting a rescue keratin construct restores CDX2 levels in K8/K18-knockdown embryos (Extended Data Fig. 7g). Therefore, keratins control the specification of the first trophectoderm cells of the embryo.

By the blastocyst stage, cells that did not inherit keratins eventually assemble a dense keratin network (Fig. 1b–d, Extended Data Fig. 1d) and display apical polarity and trophectoderm markers (Fig. 3b). This coincides with the appearance of junctional desmosomes (Extended Data Fig. 8a) and embryo cavitation, a process that requires mechanical stability to support rising intercellular pressure[39]. Although K8/K18-knockdown embryos still cavitate and form a blastocyst, they exhibit decreased volume, higher junctional tortuosity, greater surface curvature indicative of lower apical tension, and reduced cytoplasmic stiffness, assessed by tracking the movement of cytoplasmic nanoparticles[40] (Extended Data Fig. 8b–e). These defects are reversed using rescue keratin constructs (Extended Data Fig. 8f). Thus, in addition to specifying the first trophectoderm cells, keratins subsequently confer mechanical support for blastocyst morphogenesis.

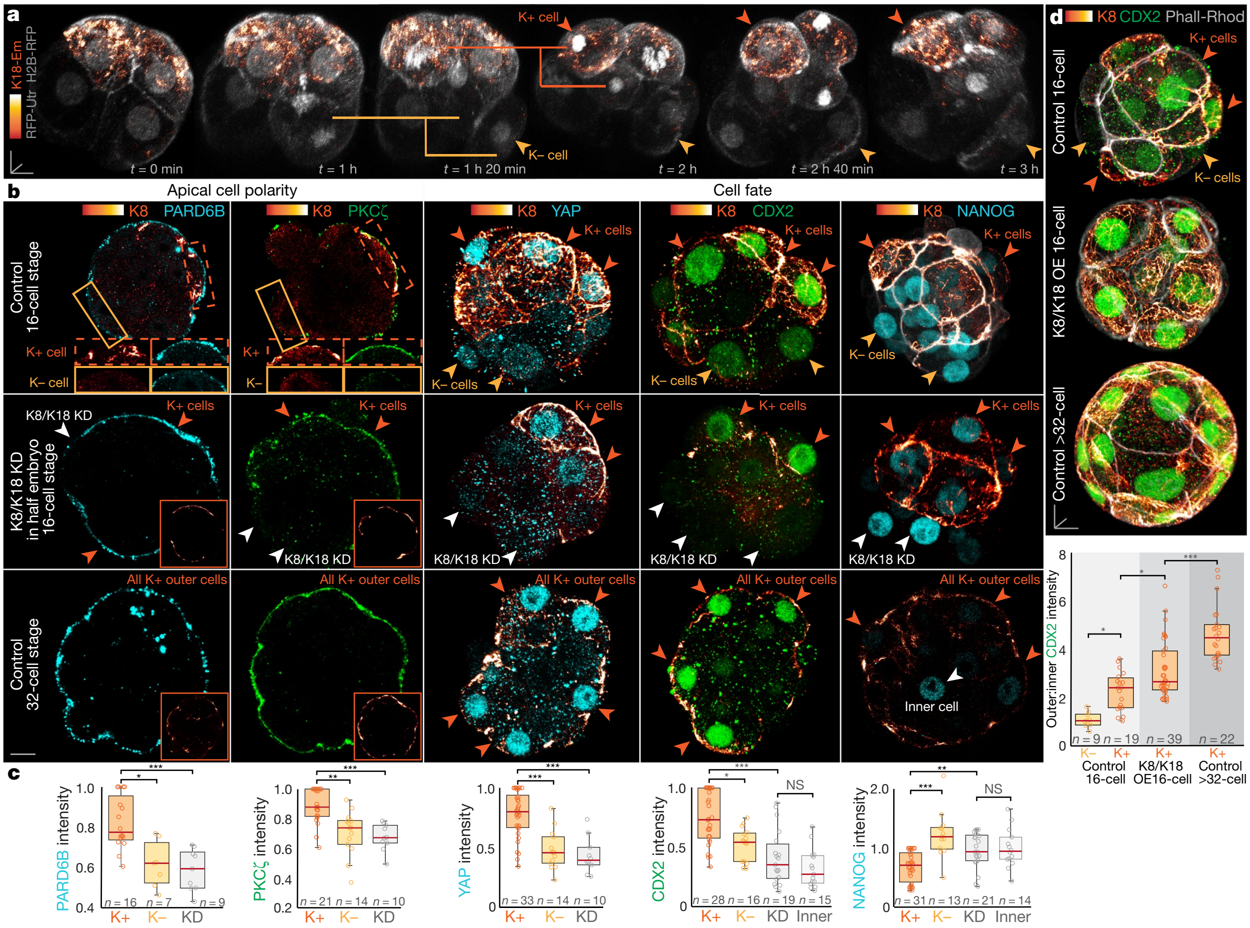


Fig. 3 | Keratin inheritance specifies the first trophectoderm cells of the embryo. a, Live-embryo imaging shows that outer cells inheriting keratins (K+) establish an extensive network after division, whereas those that did not remain devoid of filaments (K–). Data are from five independent experiments. **b**, Immunofluorescence in non-injected embryos shows that K+ cells are the first to display high levels of apical polarity and trophectoderm fate markers (top), but K8/K18-knockdown cells fail to establish these features (middle). By the 32-cell stage, the remaining cells of the embryo establish a keratin network and trophectoderm identity (bottom). Data are from three independent experiments. **c**, Quantification of fluorescence intensities. For PARD6B, $*P = 0.01$; $***P = 0.0006$. For PKCζ, $**P = 0.001$; $***P = 0.0002$. For YAP, $***P < 0.0001$. For CDX2, $*P = 0.01$; $***P < 0.0001$. For NANOG, $***P < 0.0001$; $**P = 0.004$; Kruskal–Wallis test for PARD6B, PKCζ and CDX2; ANOVA test for YAP and NANOG. **d**, Keratin overexpression (K8/K18 OE) causes premature establishment of a keratin network and trophectoderm fate throughout the 16-cell embryo. $*P = 0.03$ for control 16-cell; $*P = 0.02$ for K8/K18 OE; $***P < 0.0001$; ANOVA test. Data are from three independent experiments. In box plots, the centre line is the median, box edges show upper and lower quartiles and whiskers represent the range. Scale bars, 10 µm.

As keratins first appear in a subset of cells and function as fate determinants, they could link heterogeneities within the early embryo to lineage specification[1]. At the eight-cell stage, keratin-forming cells are connected by a microtubule bridge that links sister cells[41], indicating that they originate from the same four-cell blastomere (Fig. 4a, Extended Data Fig. 9a). Hence, we assessed whether they derive from the vegetal blastomere of the four-cell embryo (Extended Data Fig. 9b), shown to produce more trophectoderm than ICM progeny[4,42]. Selective photoactivation of the vegetal blastomere followed by staining for endogenous keratins at the eight-cell stage, and imaging live embryos expressing K18-Emerald during the four- to eight-cell window demonstrate that the vegetal blastomere preferentially produces keratin filament-forming cells (Fig. 4b, c, Extended Data Fig. 9c, d).

We finally focused on the BRG1-associated factor (BAF) chromatin remodelling complex, which promotes trophectoderm differentiation and is negatively regulated by the histone methyltransferase CARM1 that biases ICM fate[43,44]. The vegetal blastomere has the highest levels of BAF155, the main regulatory component of the BAF complex[43] (Fig. 4d). Higher BAF155 expression is also maintained in the first eight-cell blastomeres that assemble keratin filaments (Fig. 4e). Thus, we tested whether establishing four-cell embryos with different BAF patterns alters keratin expression at the eight-cell stage, by microinjecting *BAF155* (also known as *Smarcc1*) siRNAs or high levels of *BAF155* mRNA into one-cell embryos, or into one cell of two-cell embryos (Fig. 4f, Extended Data Fig. 9e). This generates patterns ranging from no detectable BAF155, to higher than normal BAF155 in all blastomeres. BAF155 knockdown or overexpression within all blastomeres triggers a reduction or increase in keratin expression, respectively. Consistently, when BAF155 levels are manipulated in half of the embryo, the resulting embryos display greater variability in keratin expression. To further determine how BAF155 regulates keratins, we used the transcriptional inhibitor actinomycin D, which eliminates keratin expression (Extended Data Fig. 9f). By contrast, facilitating transcription using trichostatin A elicits widespread keratin expression in most cells, and bypassing keratin transcription by microinjection of *K8* and *K18* mRNAs induces premature and extensive keratin expression (Extended Data Fig. 9f, g), which suggests that keratin expression is transcriptionally regulated. Furthermore, embryos that overexpress BAF155 no longer display high keratin levels when treated with actinomycin D (Extended Data Fig. 9h).

CARM1 methylates BAF155 at residue R1064[44] and CARM1-knockout embryos display lower levels of BAF155 methylation[43]. Consistently, the vegetal blastomere not only has the lowest CARM1[4] and highest

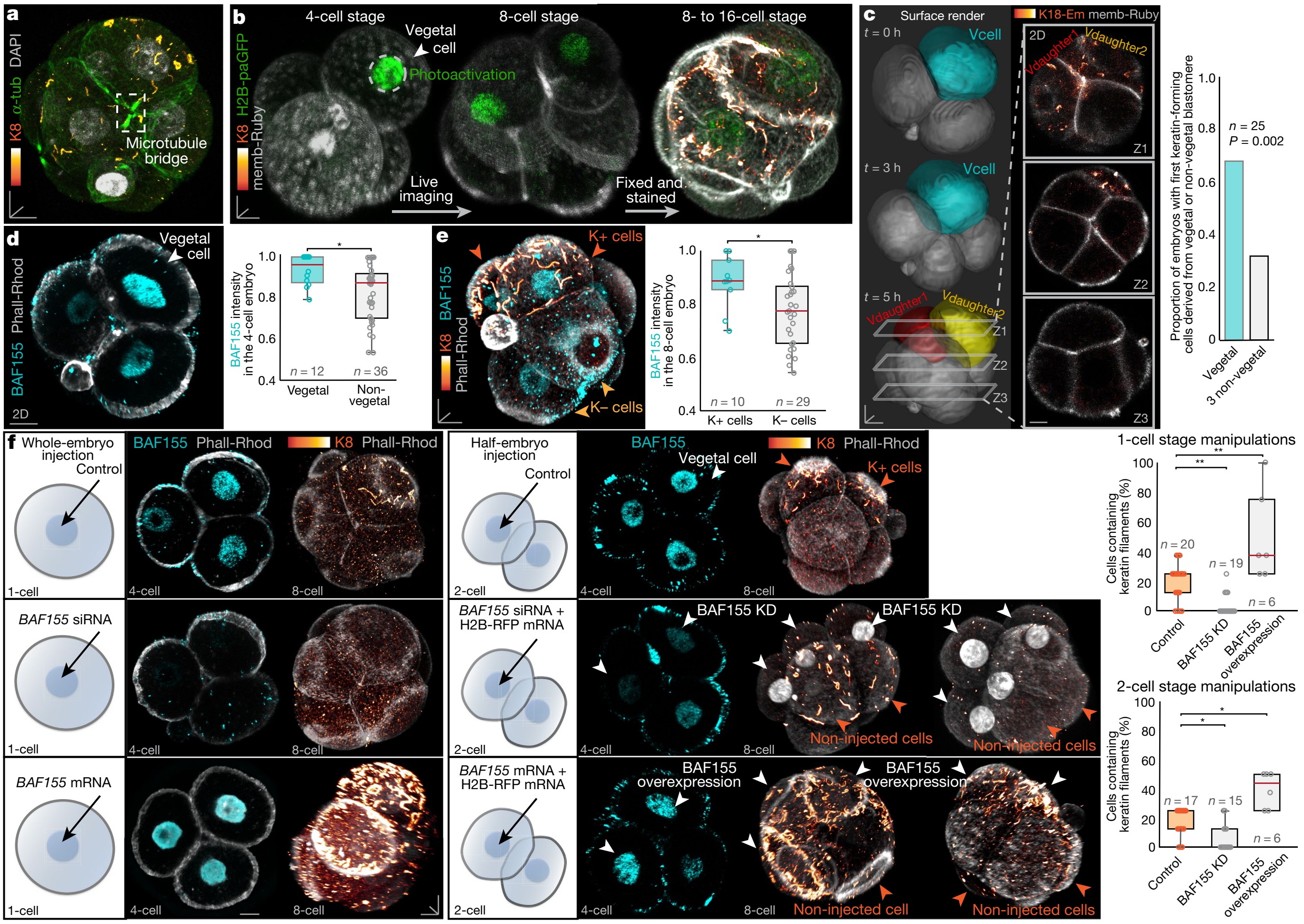


Fig. 4 | Keratin expression is regulated by early heterogeneities in the BAF complex. a, A microtubule bridge connecting sister cells reveals that the first eight-cell blastomeres assembling keratins originate from a common four-cell blastomere. Data are from three independent experiments. **b**, Selective H2B-paGFP photoactivation marks the vegetal blastomere that then produces the first keratin-forming cells. Data are from three independent experiments. **c**, Live-imaging of K18-Emerald during the four- to eight-cell stage confirms that the vegetal blastomere produces the first keratin-forming cells. *Z*-slices show keratin filaments in the cells derived from the vegetal blastomere. Graph shows proportion of embryos in which keratin-forming cells derive from vegetal blastomeres. $^{**}P = 0.002$, χ^2 test. **d**, The vegetal blastomere displays the highest endogenous BAF155 levels. $^{*}P = 0.01$; Mann–Whitney *U*-test. **e**, The first cells to form keratin filaments in the 8-cell embryo express higher BAF155 levels than cells without keratins. $^{*}P = 0.03$, Mann–Whitney *U*-test. **f**, Experimental manipulation of BAF155 levels produce different patterns of keratin expression. BAF155 knockdown reduces the proportion of cells expressing keratins at the eight-cell stage, whereas BAF155 overexpression can induce ectopic keratin expression. $^{*}P = 0.03$, $^{**}P = 0.005$ for KD; $^{**}P = 0.004$ for overexpression; ANOVA test for one-cell manipulations; Kruskal–Wallis test for two-cell manipulations. In box plots, the centre line is the median, box edges show upper and lower quartiles and whiskers represent the range. Scale bars, 10 µm.

total BAF155 levels (Fig. 4d), but also the lowest levels of methylated BAF155 (Extended Data Fig. 9i). Overexpression of CARM1 disrupts keratin expression, similarly to BAF155 knockdown, and overexpression of a BAF155(R1064K) mutant that cannot be methylated by CARM1[44] causes premature keratin expression (Extended Data Fig. 9j–l). Finally, CDX2 expression is diminished in 16-cell embryos after BAF155 knockdown or CARM1 overexpression (Extended Data Fig. 9m, n). Thus, CARM1 methylation of BAF155 leads to the differential regulation of keratins.

In conclusion, keratins function as asymmetrically inherited factors that specify the first trophectoderm cells of the embryo (Extended Data Fig. 10). Our findings validate a key aspect of the 'cell polarity' model[11] by identifying keratins as an asymmetrically inherited fate determinant. Yet, they also highlight important distinctions by showing that eight-cell blastomeres are not equivalent. Although all cells initially display apical domains, only a subset expresses keratins. Therefore, even before inner–outer segregation, cells acquire differences in cytoskeletal organization biasing their fate. Moreover, differential expression of keratins is traced back to BAF heterogeneities within the four-cell embryo, providing a mechanism to understand how early cell-to-cell variability is transmitted through divisions to influence lineage fate. This extends the idea that the fate of early blastomeres is predictable[4–6,8,10,43,45–47].

Our study also reveals interactions between the actin cortex and keratins that are important for trophectoderm specification. The apical domain first promotes apical enrichment of keratins, but after division, keratins stabilize the cortex to prevent cell internalization, support apical polarization and promote acquisition of the first hallmarks of trophectoderm specification. At later stages, CDX2 was shown to promote keratin expression[27]. Thus, the initial effect of keratins in promoting CDX2 expression could feedback into the production of more keratins to support the expansion of the keratin network for blastocyst morphogenesis. Finally, the comparable cell-to-cell variability and localization of keratins in the human embryo suggest that keratin asymmetric inheritance may represent a conserved mechanism of lineage specification in early mammalian development.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2647-4.

1. White, M. D., Zenker, J., Bissiere, S. & Plachta, N. Instructions for assembling the early mammalian embryo. *Dev. Cell* **45**, 667–679 (2018).
2. Knoblich, J. A. Asymmetric cell division: recent developments and their implications for tumour biology. *Nat. Rev. Mol. Cell Biol*. **11**, 849–860 (2010).
3. Tarkowski, A. K. & Wróblewska, J. Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. *J. Embryol. Exp. Morphol*. **18**, 155–180 (1967).
4. Torres-Padilla, M. E., Parfitt, D. E., Kouzarides, T. & Zernicka-Goetz, M. Histone arginine methylation regulates pluripotency in the early mouse embryo. *Nature* **445**, 214–218 (2007).
5. White, M. D. et al. Long-lived binding of Sox2 to DNA predicts cell fate in the four-cell mouse embryo. *Cell* **165**, 75–87 (2016).
6. Wang, J. et al. Asymmetric expression of LincGET biases cell fate in two-cell mouse embryos. *Cell* **175**, 1887–1901.e18 (2018).
7. Biase, F. H., Cao, X. & Zhong, S. Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. *Genome Res*. **24**, 1787–1796 (2014).
8. Goolam, M. et al. Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse rmbryos. *Cell* **165**, 61–74 (2016).
9. Shi, J. et al. Dynamic transcriptional symmetry-breaking in pre-implantation mammalian embryo development revealed by single-cell RNA-seq. *Development* **142**, 3468–3477 (2015).
10. Casser, E. et al. Totipotency segregates between the sister blastomeres of two-cell stage mouse embryos. *Sci. Rep*. **7**, 8299 (2017).
11. Johnson, M. H. & Ziomek, C. A. The foundation of two distinct cell lineages within the mouse morula. *Cell* **24**, 71–80 (1981).
12. Maître, J. L. et al. Asymmetric division of contractile domains couples cell positioning and fate specification. *Nature* **536**, 344–348 (2016).
13. Anani, S., Bhat, S., Honma-Yamanaka, N., Krawchuk, D. & Yamanaka, Y. Initiation of Hippo signaling is linked to polarity rather than to cell position in the pre-implantation mouse embryo. *Development* **141**, 2813–2824 (2014).
14. Korotkevich, E. et al. The Apical domain is required and sufficient for the first lineage segregation in the mouse embryo. *Dev. Cell* **40**, 235–247.e7 (2017).
15. Nishioka, N. et al. The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev. Cell* **16**, 398–410 (2009).
16. Zenker, J. et al. Expanding actin rings zipper the mouse embryo for blastocyst formation. *Cell* **173**, 776–791.e17 (2018).
17. Coulombe, P. A. & Wong, P. Cytoplasmic intermediate filaments revealed as dynamic and multipurpose scaffolds. *Nat. Cell Biol*. **6**, 699–706 (2004).
18. Jackson, B. W. et al. Formation of cytoskeletal elements during mouse embryogenesis. Intermediate filaments of the cytokeratin type and desmosomes in preimplantation embryos. *Differentiation* **17**, 161–179 (1980).
19. Paulin, D., Babinet, C., Weber, K. & Osborn, M. Antibodies as probes of cellular differentiation and cytoskeletal organization in the mouse blastocyst. *Exp. Cell Res*. **130**, 297–304 (1980).
20. Oshima, R. G., Howe, W. E., Klier, F. G., Adamson, E. D. & Shevinsky, L. H. Intermediate filament protein synthesis in preimplantation murine embryos. *Dev. Biol*. **99**, 447–455 (1983).
21. Hesse, M., Franz, T., Tamai, Y., Taketo, M. M. & Magin, T. M. Targeted deletion of keratins 18 and 19 leads to trophoblast fragility and early embryonic lethality. *EMBO J*. **19**, 5060–5070 (2000).
22. Tamai, Y. et al. Cytokeratins 8 and 19 in the mouse placental development. *J. Cell Biol*. **151**, 563–572 (2000).
23. Baribault, H., Price, J., Miyai, K. & Oshima, R. G. Mid-gestational lethality in mice lacking keratin 8. *Genes Dev*. **7** (7A), 1191–1202 (1993).
24. Lu, H., Hesse, M., Peters, B. & Magin, T. M. Type II keratins precede type I keratins during early embryonic development. *Eur. J. Cell Biol*. **84**, 709–718 (2005).
25. Chisholm, J. C. & Houliston, E. Cytokeratin filament assembly in the preimplantation mouse embryo. *Development* **101**, 565–582 (1987).
26. Emerson, J. A. Disruption of the cytokeratin filament network in the preimplantation mouse embryo. *Development* **104**, 219–234 (1988).
27. Ralston, A. & Rossant, J. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Dev. Biol*. **313**, 614–629 (2008).
28. Yoon, K. H. et al. Insights into the dynamic properties of keratin intermediate filaments in living epithelial cells. *J. Cell Biol*. **153**, 503–516 (2001).
29. Fleming, T. P., Garrod, D. R. & Elsmore, A. J. Desmosome biogenesis in the mouse preimplantation embryo. *Development* **112**, 527–539 (1991).
30. Den, Z., Cheng, X., Merched-Sauvage, M. & Koch, P. J. Desmocollin 3 is required for pre-implantation development of the mouse embryo. *J. Cell Sci*. **119**, 482–489 (2006).
31. Alarcon, V. B. Cell polarity regulator PARD6B is essential for trophectoderm formation in the preimplantation mouse embryo. *Biol. Reprod*. **83**, 347–358 (2010).
32. Doi, M. & Edwards, S. F. *The Theory of Polymer Dynamics* Vol. 73 (Oxford Univ, Press, 1988).
33. Almonacid, M., Terret, M. É. & Verlhac, M. H. Actin-based spindle positioning: new insights from female gametes. *J. Cell Sci*. **127**, 477–483 (2014).
34. Käs, J., Strey, H. & Sackmann, E. Direct imaging of reptation for semiflexible actin filaments. *Nature* **368**, 226–229 (1994).
35. Samarage, C. R. et al. Cortical tension allocates the first inner cells of the mammalian embryo. *Dev. Cell* **34**, 435–447 (2015).
36. Hirate, Y. et al. Polarity-dependent distribution of angiomotin localizes Hippo signaling in preimplantation embryos. *Curr. Biol*. **23**, 1181–1194 (2013).
37. Leung, C. Y. & Zernicka-Goetz, M. Angiomotin prevents pluripotent lineage differentiation in mouse embryos via Hippo pathway-dependent and -independent mechanisms. *Nat. Commun*. **4**, 2251 (2013).
38. Magin, T. M. et al. Lessons from keratin 18 knockout mice: formation of novel keratin filaments, secondary loss of keratin 7 and accumulation of liver-specific keratin 8-positive aggregates. *J. Cell Biol*. **140**, 1441–1451 (1998).
39. Leonavicius, K. et al. Mechanics of mouse blastocyst hatching revealed by a hydrogel-based microdeformation assay. *Proc. Natl Acad. Sci. USA* **115**, 10375–10380 (2018).
40. Wirtz, D. Particle-tracking microrheology of living cells: principles and applications. *Annu. Rev. Biophys*. **38**, 301–326 (2009).
41. Zenker, J. et al. A microtubule-organizing center directing intracellular transport in the early mouse embryo. *Science* **357**, 925–928 (2017).
42. Piotrowska-Nitsche, K., Perea-Gomez, A., Haraguchi, S. & Zernicka-Goetz, M. Four-cell stage mouse blastomeres have different developmental properties. *Development* **132**, 479–490 (2005).
43. Panamarova, M. et al. The BAF chromatin remodelling complex is an epigenetic regulator of lineage specification in the early mouse embryo. *Development* **143**, 1271–1283 (2016).
44. Wang, L. et al. CARM1 methylates chromatin remodeling factor BAF155 to enhance tumor progression and metastasis. *Cancer Cell* **30**, 179–180 (2016).
45. Plachta, N., Bollenbach, T., Pease, S., Fraser, S. E. & Pantazis, P. Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nat. Cell Biol*. **13**, 117–123 (2011).
46. Kaur, G. et al. Probing transcription factor diffusion dynamics in the living mammalian embryo with photoactivatable fluorescence correlation spectroscopy. *Nat. Commun*. **4**, 1637 (2013).
47. Tabansky, I. et al. Developmental bias in cleavage-stage mouse blastomeres. *Curr. Biol*. **23**, 21–31 (2013).

Methods

Mouse embryo work

Mouse embryo experimentation was approved by the Biological Resource Center Institutional Animal Care and Use Committee (IACUC), Agency for Science, Technology and Research (IACUC Protocol 181370). C57BL/6 wild-type 3–4-week-old female mice were superovulated using 5 IU of pregnant mare serum (PMS, National Hormone and Peptide Program) gonadotropin given intraperitoneally and 5 IU of recombinant chorionic gonadotrophin (CG, National Hormone and Peptide Program) given 48 h after and immediately before mating, according to animal ethics guidelines of the Agency for Science, Technology and Research, Singapore. Embryos were flushed from oviducts of plugged females using M2 medium (Merck) and cultured in KSOM+AA (Merck) covered by mineral oil (Sigma), at 37 °C and 5% CO$_2$. Microinjections were performed using a FemtoJet (Eppendorf). mRNA synthesis was performed on linearized plasmids using the mMESSAGE mMACHINE SP6 kit (Ambion), and purified using the RNAeasy kit (QIAGEN). For live imaging experiments, mRNAs diluted in injection buffer (5 mM Tris, 5 mM NaCl, 0.1 mM EDTA) were microinjected as follows: K8-Emerald and K18-Emerald at 150 ng µl^{-1}; mRuby2-Actin at 100 ng µl^{-1}; RFP-Utrophin at 70 ng µl^{-1}; RFP-MAP2c at 80 ng ul^{-1}; memb-mRuby2 at 70 ng µl^{-1}; H2B-RFP and H2B-GFP at 5 ng µl^{-1}; Desmoglein2-Emerald and Desmoglein2-mRuby2 at 150 ng µl^{-1}; H2B-paGFP at 20 ng ul^{-1}. For overexpression experiments, mRNAs were microinjected as follows: K8 and K18 at 300 ng µl^{-1}; BAF155 and BAF155(R1064K) at 500 ng µl^{-1}; CARM1 at 300 ng µl^{-1}. siRNAs (QIAGEN) were microinjected at the following concentrations: K8 (500 nM), K18 (500 nM), DSG2 (200 nM), DSC3 (200 nM), desmoplakin (200 nM), plakoglobin (200 nM), PARD6B (200 nM), YAP1 (200 nM), BAF155 (500 nM).

The siRNAs used are: Mm_Krt2-8_1 (AACCATGTACCAGATTAAGTA), Mm_Krt2-8_2 (ATGGATGGCATCATCGCTGAA), Mm_Krt1-18_1 (CAGAGTGGTGTCCGAGACTAA), Mm_Krt1-18_3 (CCGGGAACATCTGGAGAAGAA), Mm_Dsg2_1 (CAGCATTATGCCAATGAAGAA), Mm_Dsg2_2 (CTCCGTCACTTCAGAGATTAA), Mm_Dsc3_2 (CAGAGATAATTCAAGATTATA), Mm_Dsc3_5 (AACTGCGGATGTTCAAATATA), Mm_Dsp_2 (CAGGAAGTTCTTCGATCAATA), Mm_Dsp_4 (ACCGGTTGACATGGCGTATAA), Mm_Jup_4 (CAGACAGTACACACTCAAGAA), Mm_Jup_5 (CACTATGGCTATGGCCACTAA), Mm_Pard6b_3 (CACGGGCCTGCTAGCTGTCAA), Mm_Pard6b_4 (CAGGTGACTGACATGATGATA), Mm_Yap1_6 (ACCCTTGAACATATACATTTA), Mm_Yap1_7 (AACATCCTATTTAAATCTTAA), Mm_Smarcc1_5 (ACGCATCCTGGTTTGATTATA), Mm_Smarcc1_6 (TCGAACTGACATTTACTCCAA).

For drug treatments, all drugs were diluted in KSOM to the following concentrations: cytochalasin D at 20 µg ml^{-1}, SiR-Actin at 100 nM, MG-132 at 25 µM, actinomycin D at 100 ng ml^{-1}, trichostatin A at 75 nM. Drugs were applied for 2 h before embryo fixation, with the exception of actinomycin D and trichostatin A, which were both applied for the entire 4- to 8-cell stage window to effectively block or promote transcription respectively.

Human embryo work

Human embryos were donated to the Reproductive Medicine Research Center, Sixth Affiliated Hospital of Sun Yat-sen University for research purposes, following ethical guidelines of the Sixth Affiliated Hospital of Sun Yat-sen University. Experiments were performed according to the guidelines of the Institute of Zoology, Chinese Academy of Sciences and the Sixth Affiliated Hospital of Sun Yat-sen University.

This work was approved by the Ethics Committee of Center for Reproductive Medicine, Sixth Affiliated Hospital of Sun Yat-Sen University (Research license 2019SZZX-008). The Medicine Ethics Committee of Center for Reproductive Medicine, Sixth Affiliated Hospital of Sun Yat-Sen University is composed of 11 members, including experts of laws, scientists and clinicians with relevant expertise. The Committee evaluated the scientific merit and ethical justification of this study and conducted a full review of the donations and use of these samples.

All embryo donor couples signed informed consent forms for voluntary donations of surplus embryos for research, at the Center for Reproductive Medicine, Sixth Affiliated Hospital of Sun Yat-Sen University. Participation in the study was voluntary and no financial inducements were offered for embryo donation. The culture of all embryos was terminated before day 14 post-fertilization. Couples were informed that their embryos would be used to study the developmental mechanisms of human embryos and that their donation would not affect their IVF cycle. The informed consent forms clearly state the goals of the research, clinical procedures used in the study, potential benefits and risks to research participants, and steps taken to ensure that the privacy of each embryo donor was well protected. The participation of embryo donors in the study can only be obtained if eligible participants were provided with all necessary information about the study and the opportunity to receive counselling. These informed consent guidelines are in line with the ethical and regulatory framework set forth by the Center for Reproductive Medicine, Sixth Affiliated Hospital of Sun Yat-sen University, and complied with the International Society for Stem Cell Research (ISSCR) Guidelines for Stem Cell Research and Clinical Translation (2016) and Ethical Guidelines for Human Embryonic Stem Cell Research (2003) jointly issued by the Ministry of Science and Technology and the Ministry of Health of the People's Republic of China.

All donated samples in this study were obtained from frozen embryos from couples who signed informed consent agreements. The study employed standard clinical protocols for embryo collection, cryopreservation, thawing and culture procedures. Human embryos were frozen–thawed 3 or 5 days post-fertilization. Cryopreserved embryos were thawed using Kitazato Thawing Media Kit VT802 (Kitazato Dibimed) depending on the protocol used for freezing and following the manufacturer's instructions. The embryos were cultured in Single-step embryo culture medium (LifeGlobal) covered with oil (LifeGlobal) (from 4-cell stage to blastocyst stage). Embryos with normal morphology and cleavage patterns were used in this study.

Microscopy

Imaging was performed using a laser scanning confocal microscope (LSM 780 and LSM 880, Zeiss) with a water UV-VIS-IR Apochromat 63× 1.2 NA objective. For live imaging, embryos were cultured in LabTek chambers (Nunc) in KSOM+AA (Merck) covered by mineral oil (Sigma), using the incubator system adapted for the microscope (Carl Zeiss, Jena) to maintain the embryos at 37 °C and 5% CO$_2$. Embryos were scanned every 15 to 20 min for long-term imaging, and selected mitotic cells were imaged at higher temporal resolution of 1 to 3 min intervals in order to track the dynamics of keratin filaments throughout the entire cell division. FRAP was performed at 3.5-times zoom on a 5 µm × 10 µm region of interest, photobleached using the 488 nm laser at 100%, with a pixel dwell time of 6 µs and scanning speed of 6. For photoactivation experiments, H2B-paGFP was selectively illuminated in the nuclei of vegetal blastomeres using an 820 nm two-photon laser (Mai Tai, Spectra-Physics) as described[5,45], followed by live-imaging using a 488 nm laser to track the photoactivated signal in the daughter cells.

For measurements of cell elasticity, 0.1-µm-diameter carboxylate-modified FluoSpheres (Invitrogen) were microinjected into live embryos. The size of these nanoparticles is larger than the average mesh size of the cytoskeletal network[40]. We optimized their concentration to obtain an average of 10 beads per cell, homogenously distributed throughout the cytoplasm at blastocyst stage. Tracking of the movement of these nanoparticles was performed by imaging individual particles at 30-times zoom, 50 frames s^{-1} for 2 min, as previously described[48].

Immunofluorescence

Embryos were fixed in 4% paraformaldehyde for 30 min at room temperature or overnight at 4 °C, washed twice in PBS with 0.1% Triton X-100, permeabilized for 20 min in PBS with 0.5% Triton X-100, and incubated in PBS with 10% fetal bovine serum (blocking solution) for 30 min. Embryos were then incubated at 4 °C overnight in primary antibodies diluted in blocking solution, at the following concentrations: K8 (DSHB) at 1:20, K18 (Sigma, SAB4501665) at 1:200, K19 (DSHB) at 1:50, pan-Keratin (Cell Signaling, 4545) at 1:50, α-tubulin (Sigma, T6199) at 1:1,000, PARD6B (Santa Cruz, 166405) at 1:50, PKCζ (Santa Cruz, 17781) at 1:50, AMOT (gift from H. Sasaki) at 1:200, YAP (Cell Signaling, 8418S) at 1:500, CDX2 (Abcam, 88129) at 1:200, NANOG (Abcam, 80892) at 1:200, desmoglein1/2 (Progen, 61002S) undiluted, plakoglobin (Progen, 61005S) undiluted, plakophilin (Progen, 651101S) undiluted, desmoplakin1 (Progen) at 1:100, BAF155 (Santa Cruz, 48350) at 1:50, dimethyl-BAF155 (Merck, ABE1339) at 1:100, and CARM1 (Cell Signaling, 3379S) at 1:150. After primary antibody incubation, embryos were washed 5 times for 20 min in PBS with 0.1% Triton X-100 and incubated 1.5 h at room temperature or overnight at 4 °C in secondary antibodies diluted in blocking solution to 1:500. Phalloidin-Rhodamine (Molecular Probes, R415) diluted to 1:500 and NucBlue Fixed Cell Stain ReadyProbes reagent (Invitrogen) diluted to 1:100 in blocking solution were also used to label the F-actin and chromatin respectively. Embryos were washed three times in PBS with 0.1% Triton X-100 before mounting in PBS covered with mineral oil (Sigma) in an 8-well LabTek chamber (Nunc).

Image analysis

Image analyses were performed using Imaris 8.2 (Bitplane AG), Fiji, and MATLAB. 3D segmentation of whole embryos, individual cells and individual nuclei was performed using the Imaris manual surface rendering module, and 3D segmentation of keratin filaments was done with the automatic surface rendering mode. The Imaris statistics module was used to obtain values for total fluorescence intensities, cell volumes, and nucleus volumes. Measurements of apical fluorescence intensities of PARD6B and PKCζ were performed in Fiji by selecting the apical region of individual cells and averaging the fluorescence intensities across five different Z-planes. All quantifications were normalized to background fluorescence to correct for weaker fluorescence with increasing depth through the embryo. M.B. thanks Bitplane AG for an Imaris Developer License.

To quantify the localization of keratin filaments within individual cells, we used a polarization index, adapted from previous work[49,50]. Spatial coordinates for individual keratin filaments and the cell centre of mass, volumes of individual keratin filaments, and lengths of the cell apical-basal axis were obtained using Imaris software. The polarization index was calculated by obtaining the difference between the volume-weighted average position of all keratin filaments within the cell and the position of the cell centre of mass, normalized to the cell apical–basal axis.

For calculations of keratin filament movement during interphase and mitosis, the spatial coordinates of each filament were obtained from the Imaris statistics module. The mean speed of filament movement was calculated by dividing the distance between the initial and final positions of the filament by the elapsed time.

For analysis of FRAP experiments, mean fluorescence intensity at the photobleached region of interest (ROI) was corrected by background fluorescence and normalized to a non-photobleached reference. The average of the pre-bleach fluorescence intensities was set to 100%, and the fluorescence intensity immediately after photobleaching was set to 0%. The normalized mean fluorescence intensities were then fitted with an exponential function, as previously described[16]. The immobile fraction was calculated by taking $1 - I_\infty$, in which I_∞ is the normalized mean fluorescence intensity when the intensity recovers to a plateau. All fittings were performed in MATLAB, and FRAP kymographs were created using the Montage tool in Fiji.

To characterize the morphology of cell–cell junctions, we used a tortuosity index calculated by measuring the total junction length, normalized to the Euclidean distance.

For measurements of surface curvature, we first segmented individual cells in Imaris and extracted the apical surface of each cell using a custom MATLAB code. The radius of surface curvature was then determined using the radius of the sphere that best fits the cell apical surface.

Statistical analysis

Statistical analyses were performed in GraphPad Prism and Excel. Qualitative data were represented using a contingency table and analysed using a Fisher's exact test or χ^2 test. All quantitative data were first analysed for normality using a D'Agostino-Pearson omnibus normality test. Variables showing a normal distribution were then analysed using an unpaired, two-tailed Student's *t*-test or ANOVA with Tukey's multiple comparisons test for two groups or more than two groups respectively. Variables that did not show a normal distribution were analysed using an unpaired, two-tailed Mann–Whitney *U*-test or Kruskal–Wallis test with Dunn's multiple comparisons test for two groups or more than two groups respectively. No statistical test was performed to determine sample size, and sample size was determined based on previous experience and in accordance to previous studies. Embryos were randomly allocated into experimental groups and randomly selected for analysis. Reproducibility was confirmed by at least three independent experiments. The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Source data are provided with this paper.

Code availability

Code for apical surface curvature analysis has been published in a publicly available repository at https://github.com/gracelhy/Analysis-of-embryo-parameters.

48. Daniels, B. R., Masi, B. C. & Wirtz, D. Probing single-cell micromechanics in vivo: the microrheology of C. elegans developing embryos. *Biophys. J.* **90**, 4712–4719 (2006).
49. Park, H. Y., Trcek, T., Wells, A. L., Chao, J. A. & Singer, R. H. An unbiased analysis method to quantify mRNA localization reveals its correlation with cell motility. *Cell Rep.* **1**, 179–184 (2012).
50. Skamagki, M., Wicher, K. B., Jedrusik, A., Ganguly, S. & Zernicka-Goetz, M. Asymmetric localization of Cdx2 mRNA during the first cell-fate decision in early mouse development. *Cell Rep.* **3**, 442–457 (2013).

Acknowledgements We thank X. Liang for assistance with human embryo work. This work was supported by grants from ASTAR, EMBO, and HHMI to N.P., EMBL Australia to M.B., and the ASTAR Graduate Scholarship to H.Y.G.L.

Author contributions H.Y.G.L. conceived the project, performed the experiments and data analysis, and wrote the manuscript with contributions from all other authors. Y.D.A. and M.G. assisted with experiments and data analysis. Y.W. and H.W. performed human embryo studies. P.T. and S.B. performed mouse work and embryo microinjection experiments. M.B. contributed to data analysis and manuscript writing. N.P. supervised the project.

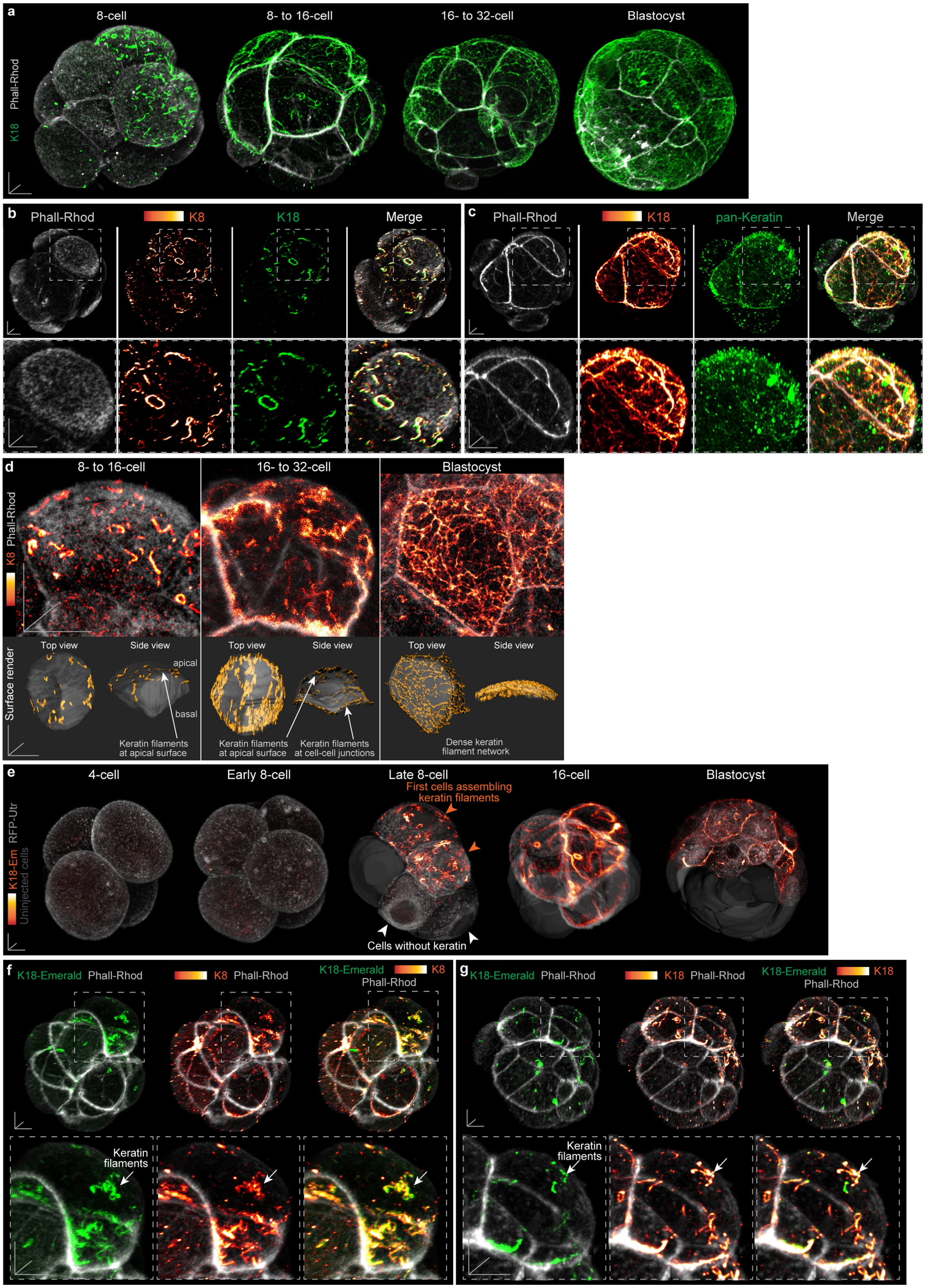
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2647-4.

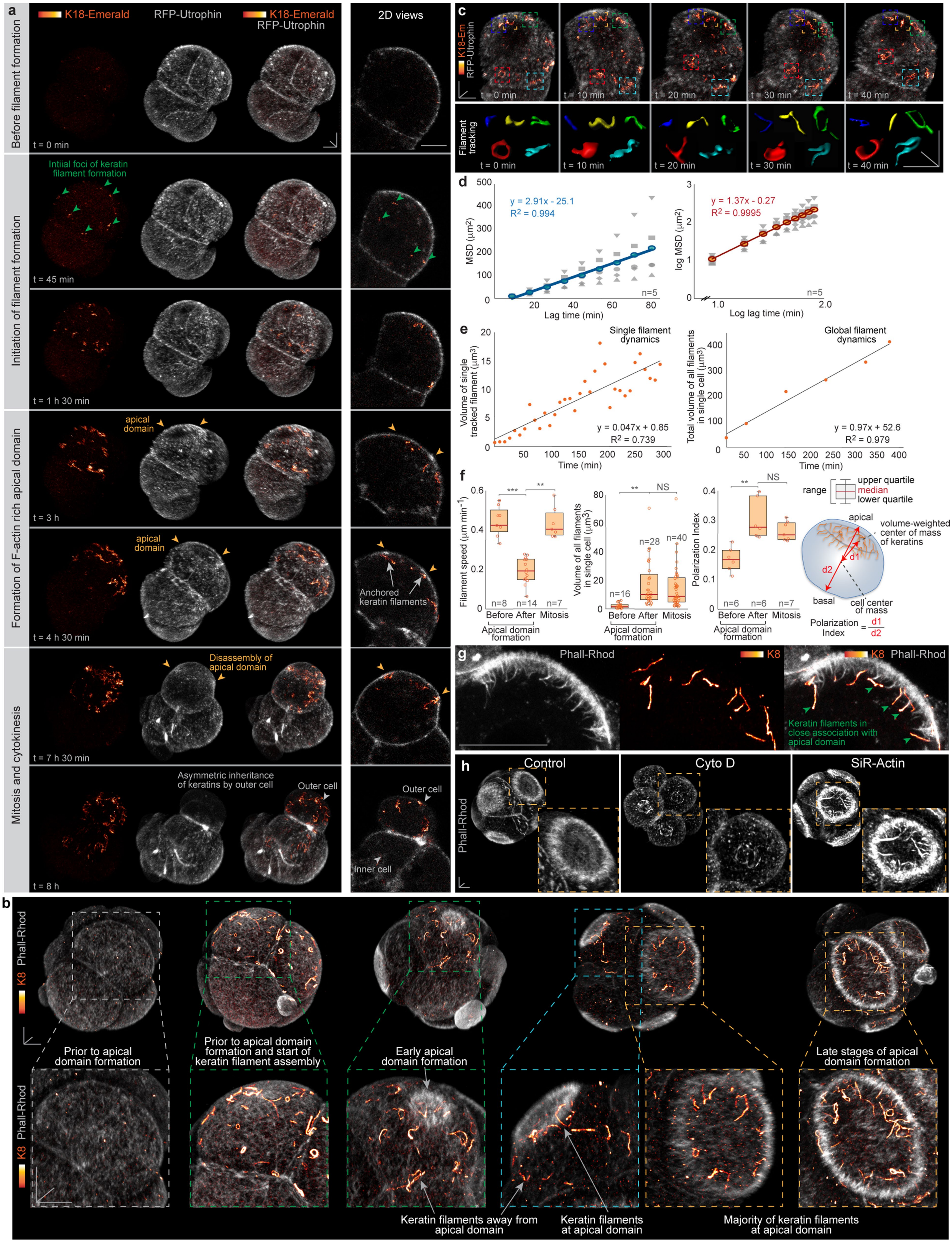
Correspondence and requests for materials should be addressed to N.P.

Peer review information *Nature* thanks Magdalena Zernicka-Goetz and the other anonymous reviewer(s) for their contribution to the peer review of this work.



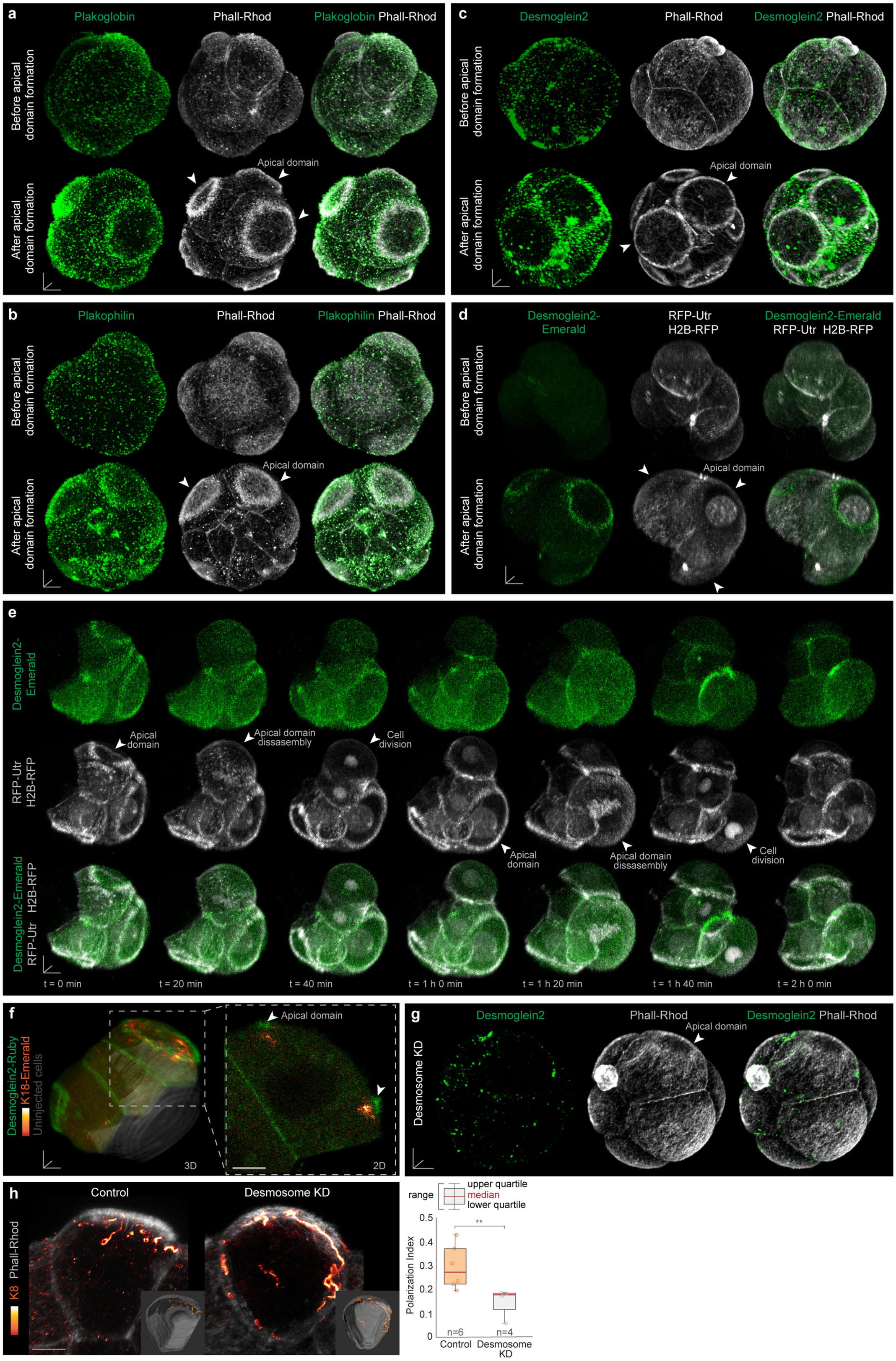
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Keratin filaments in the preimplantation mouse embryo. a, 3D views of mouse embryos at multiple developmental stages, stained for K18. K18 expression and localization resemble that of K8. Note the initial assembly of filaments in a specific subset of cells in the 8-cell embryo. Data are from five independent experiments. **b**, Double immunofluorescence for K8 and K18 shows their colocalization in filament structures within the same embryo. Data are from three independent experiments. **c**, Double immunofluorescence using a pan-keratin antibody and K18 shows colocalization in filament structures within the same embryo. Data are from three independent experiments. **d**, High-magnification views highlight keratin filament organization at multiple developmental stages (top). Surface render of computationally-segmented cells and keratin filaments with top and side views show the changes in cell morphology and keratin filament organization at different developmental stages. The density of the keratin filament network increases over time and the filaments become enriched at cell-cell junctions. Data are from five independent experiments. **e**, Live imaging of embryos expressing K18-Emerald. A subset of cells begins to assemble keratin filaments at the eight-cell stage, similar to observations from immunofluorescence for endogenous keratins. No keratin filaments are detected in four-cell or early uncompacted eight-cell embryos. Data are from three independent experiments. **f**, **g**, Colocalization of K18-Emerald and immunofluorescence against K8 (**f**) or K18 (**g**). Bottom panels show zoomed views of single cells expressing keratin filaments, with arrows pointing to an example of signal colocalization. Data are from three independent experiments. Scale bars, 10 µm.



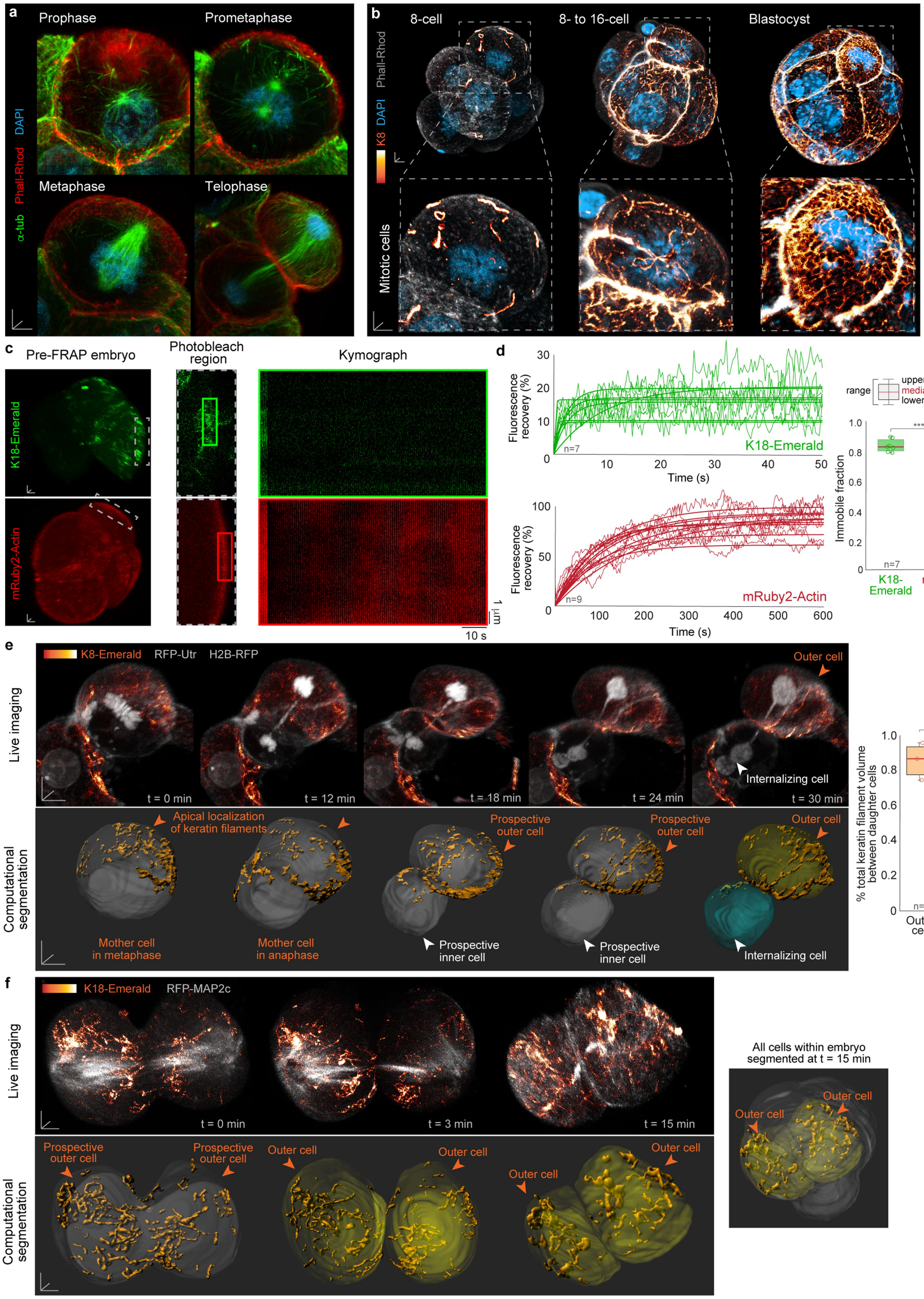
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Tracking and quantitative analysis of keratin filament movement during interphase and mitosis. a, Time series of an embryo expressing K18-Emerald and RFP-Utrophin, with the corresponding major cellular events labelled in the left column. Separate K18-Emerald and RFP-Utrophin channels are shown. Right panels show 2D views through a single cell that assembles keratin filaments, for better visualization of keratin distribution within the cell, relative to the apical domain. Keratin filament assembly is initiated before the formation of the apical domain. When the apical domain forms, keratin filaments become enriched apically in close association with F-actin. During mitosis, the apical domain disassembles but keratin filaments remain apically localized, resulting in their asymmetric inheritance by the outer daughter cell. Data are from three independent experiments. **b**, Immunofluorescence of endogenous keratins in embryos fixed at different stages of apical domain formation recapitulates the pattern and localization of keratin filaments relative to the apical domain observed in live imaging experiments. Data are from three independent experiments. **c**, Computationally-rendered filaments obtained from live imaging data. In this example, five individual filaments were tracked over time with a 10-min interval between frames. Data are from three independent experiments. **d**, The log mean square displacement (MSD) versus log lag time graph indicates that the movement of keratin filaments is unconfined and diffusive (slope > 1). Pearson's correlation. **e**, Volume of an individual keratin filament and total filament volume within a single tracked cell increase linearly over time. Pearson's correlation. **f**, Quantification of filament speed, volume of filaments, and polarization index before apical domain formation, after apical domain formation, and during mitosis. After the formation of the apical domain, keratin filaments move more slowly, display a larger total volume, and become more apically polarized than before apical domain formation. During mitosis, keratin filaments move faster, but retain a large volume and high apical polarization. $^{***}P = 0.0002$; $^{**}P = 0.001$; Kruskal–Wallis test for filament speed; $^{**}P = 0.003$; ANOVA test for filament volume; $^{**}P = 0.003$; Kruskal–Wallis test for polarization index. Scheme shows the parameters used for calculation of the polarization index. $d1$ is the distance between the volume-weighted centre of mass of the keratin filaments and the centre of mass of the cell. $d2$ is the length of the apical-basal axis of the cell. **g**, High-resolution immunofluorescence images show that keratin filaments align specifically along actin filaments extending from the apical domain. Green arrows indicate examples of keratin–actin colocalization. Data are from three independent experiments. **h**, Differences in F-actin accumulation at the apical domain of control embryos and embryos treated with cytochalasin D or a high concentration of SiR-Actin. Insets show zoomed views of individual 8-cell blastomeres, highlighting the loss of the apical domain in cytochalasin D-treated embryos, and a dense accumulation of apical F-actin in SiR-Actin-treated embryos. Data are from three independent experiments. Scale bars, 10 µm.



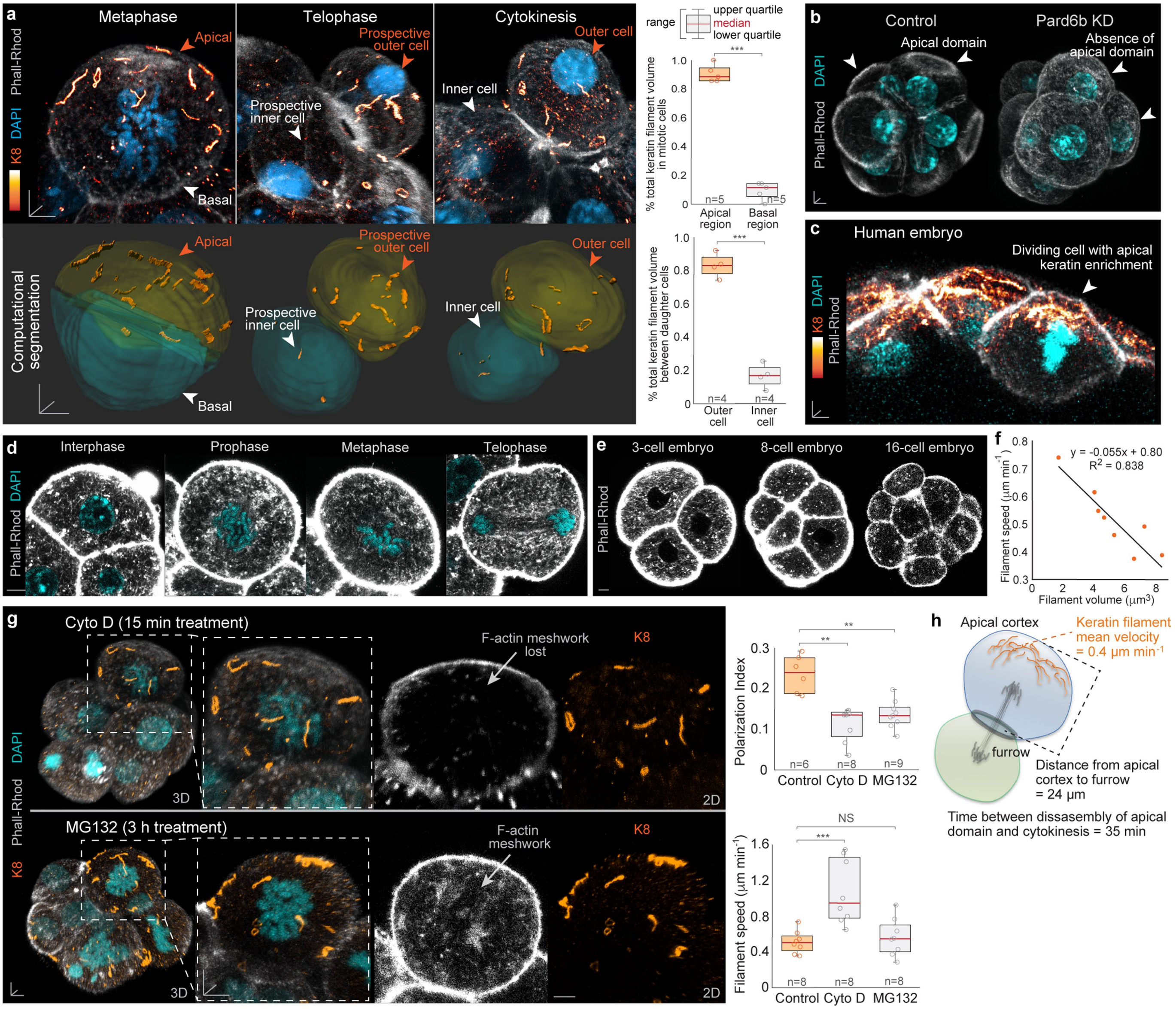
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Apical keratin localization requires desmosome protein components. **a**–**c**, Immunofluorescence of endogenous plakoglobin (**a**), plakophilin (**b**), and desmoglein2 (**c**) before and after apical domain formation. Apical accumulation of all three desmosome components is observed after apical domain formation. **d**, Live imaging of an embryo expressing desmoglein2-Emerald, RFP-utrophin and H2B-RFP recapitulates the endogenous desmoglein2 expression, both before and after apical domain formation. **e**, Time series of embryo expressing desmoglein2-Emerald, RFP-utrophin and H2B-RFP. Desmoglein2-Emerald accumulates with the apical domain (labelled by RFP-utrophin) during interphase. When the cell enters mitosis, desmoglein2-Emerald disassembles from the apical surface together with the apical domain. White arrows indicate two different mitotic events within the same embryo. **f**, Live embryo expressing desmoglein2-Ruby and K18-Emerald shows the enrichment of keratin filaments at the site of apical desmosome accumulation. **g**, Embryos injected with siRNAs against desmosome components do not accumulate desmoglein2 apically with the apical domain. **h**, Desmosome knockdown causes a more homogenous distribution of keratin filaments, as measured by a polarization index. $^{**}P = 0.01$; unpaired, two-tailed Mann–Whitney *U*-test. Data are from three independent experiments. Scale bars, 10 µm.

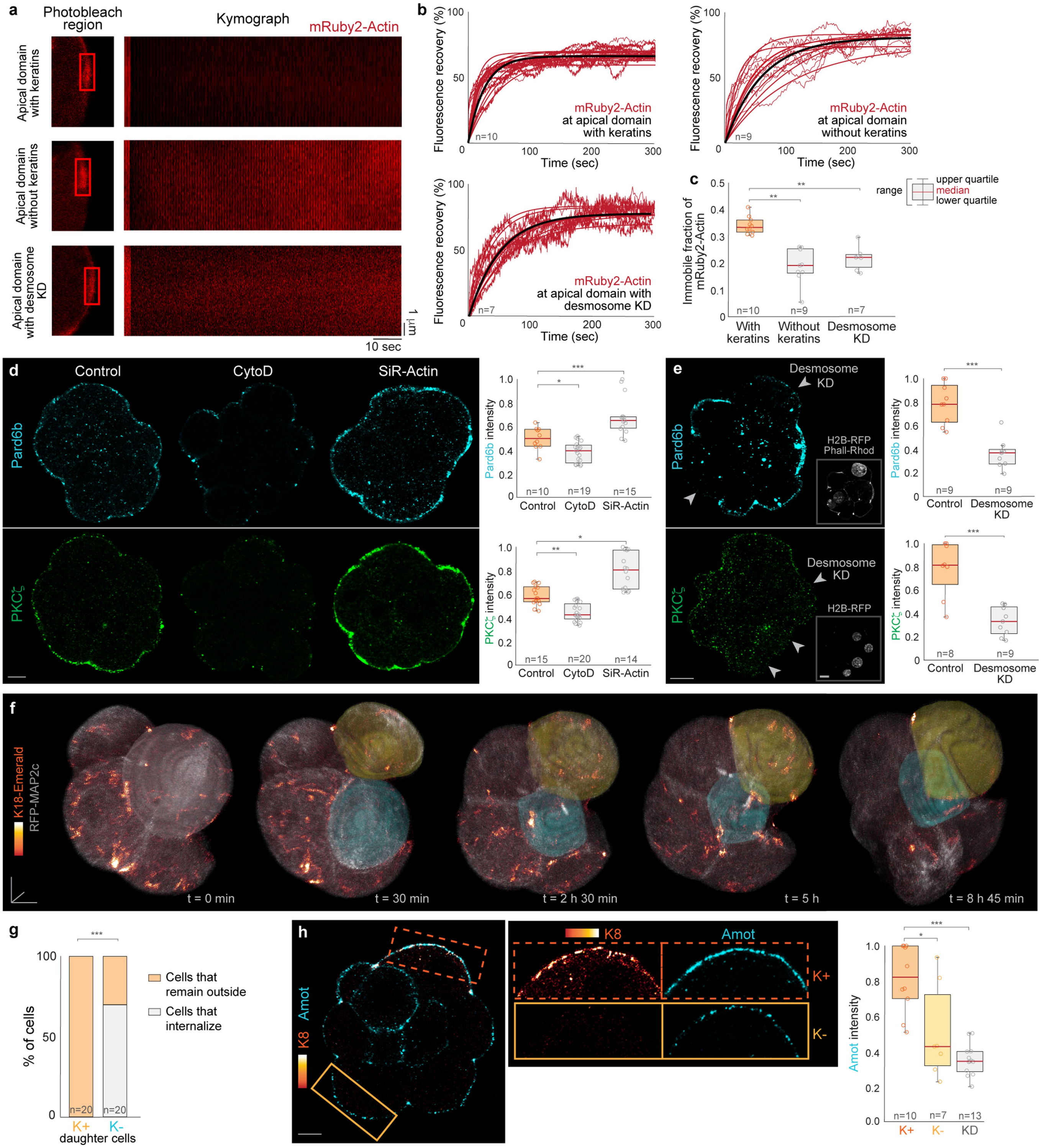


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Keratin filaments are stably retained during mitosis, and become asymmetrically inherited by outer daughter cells. a, Immunofluorescence shows the extensive remodelling of cortical F-actin and microtubules during different stages of mitosis. Data are from six independent experiments. **b**, Immunofluorescence for K8 shows endogenous keratin filaments retained within mitotic cells in embryos fixed at multiple stages of development. Data are from six independent experiments. **c**, FRAP experiments for K18-Emerald and mRuby2-Actin performed in whole live embryos. All cells selected for FRAP were at the eight-cell stage and in interphase, when the actin ring is visible. 3D views of entire pre-FRAP embryos (left), zoomed views of the photobleached regions of interest (middle), and kymographs of pre- and post-FRAP fluorescence intensities (right). Data are from three independent experiments. **d**, Analysis of FRAP experiments. Left graphs show fluorescence recovery of K18-Emerald (green) and mRuby2-Actin (red) over time. Thinner lines represent raw data after normalization, and thicker lines indicate fitted exponential curves. Right graph shows that K18-Emerald has a larger immobile fraction than mRuby2-Actin. $^{***}P < 0.0001$; unpaired, two-tailed Student's *t*-test. **e**, Live imaging of embryos expressing K8-Emerald show a similar pattern of expression and inheritance as K18-Emerald. The outer daughter cell inherits most of the keratin filaments during an outer-inner division (top). Computational segmentation of the same cell at each stage of mitosis (bottom). Quantification of proportion of keratin filaments inherited by outer and inner cells in live embryos expressing K8-Emerald shows a comparable asymmetry in keratin inheritance as K18-Emerald. $^{**}P = 0.001$; unpaired, two-tailed Student's *t*-test. **f**, Time series of a cell expressing K18-Emerald undergoing a symmetric outer-outer division. Keratin filaments are uniformly inherited by both daughter cells during divisions producing two outer cells (top). Computational segmentation of the same cells at each time point (bottom), and whole embryo inset highlighting the outer location of both daughter cells (right). Data are from four independent experiments. Scale bars, 5 µm.

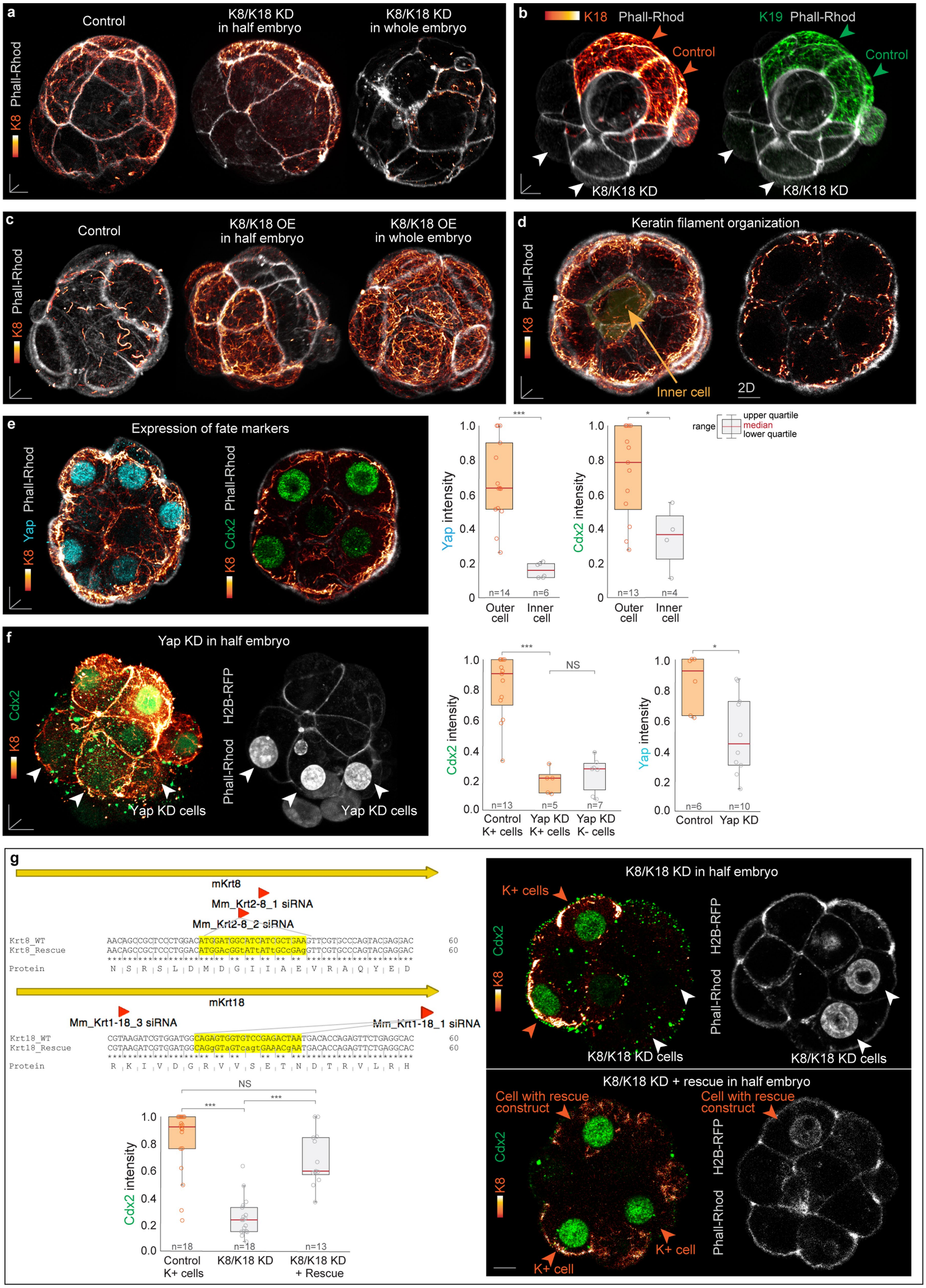


Extended Data Fig. 5 | A dense F-actin meshwork within mitotic cells hinders the movement of keratin filaments away from the apical cortex. **a**, Immunofluorescence of embryos fixed specifically when a cell was undergoing mitosis or cytokinesis. Top, keratin filaments remain apically-localized throughout different mitotic stages, and become inherited by the prospective outer cell. Bottom, computational segmentation of the same cells highlighting the apical keratin distribution and asymmetric keratin inheritance. Quantification of proportion of endogenous keratin filaments present in the apical and basal regions of mitotic cells, and between prospective outer and inner daughter cells, showing a comparable asymmetry in endogenous keratin localization and inheritance as K18-Emerald dynamics in live embryos. $***P < 0.0001$; unpaired, two-tailed Student's *t*-test. **b**, Embryos microinjected with *Pard6b* siRNAs do not form an apical F-actin ring in the eight-cell embryo. Data are from three independent experiments. **c**, Mitotic cell within a fixed human embryo also displays an apical localization of keratins. **d**, A dense cytoplasmic F-actin meshwork is maintained throughout interphase and all stages of mitosis. Data are from three independent experiments. **e**, The F-actin meshwork is also present in cells across different stages of development. Representative images of a 3-cell, compacted 8-cell, and 16-cell embryo with all cells displaying a dense cytoplasmic F-actin meshwork. Data are from three independent experiments. **f**, Analysis of keratin filament movement during mitosis reveals that filament speed is inversely related to filament volume. $n = 8$ filaments; Pearson's correlation. **g**, Acute cytochalasin D treatment for 15 min specifically during mitosis disrupts the F-actin meshwork, reduces the apical localization of keratins, and increases keratin filament speed. Cells treated with MG132 for 3 h retain an F-actin meshwork, but keratin apical localization is reduced and filament speed is unchanged. $**P = 0.002$ for CytoD; $**P = 0.01$ for MG132; Kruskal–Wallis test for polarization index; $***P = 0.0005$; ANOVA test for filament speed. **h**, Scheme of a cell division producing an inner (green) and an outer (blue) cell. Keratin filaments localize close to the apical cortex of the forming outer daughter cell. The distance between the apical cortex and cytokinetic furrow, time between disassembly of the apical F-actin domain and cytokinesis, and the mean speed of keratin filament movement are indicated. Scale bars, 5 µm.



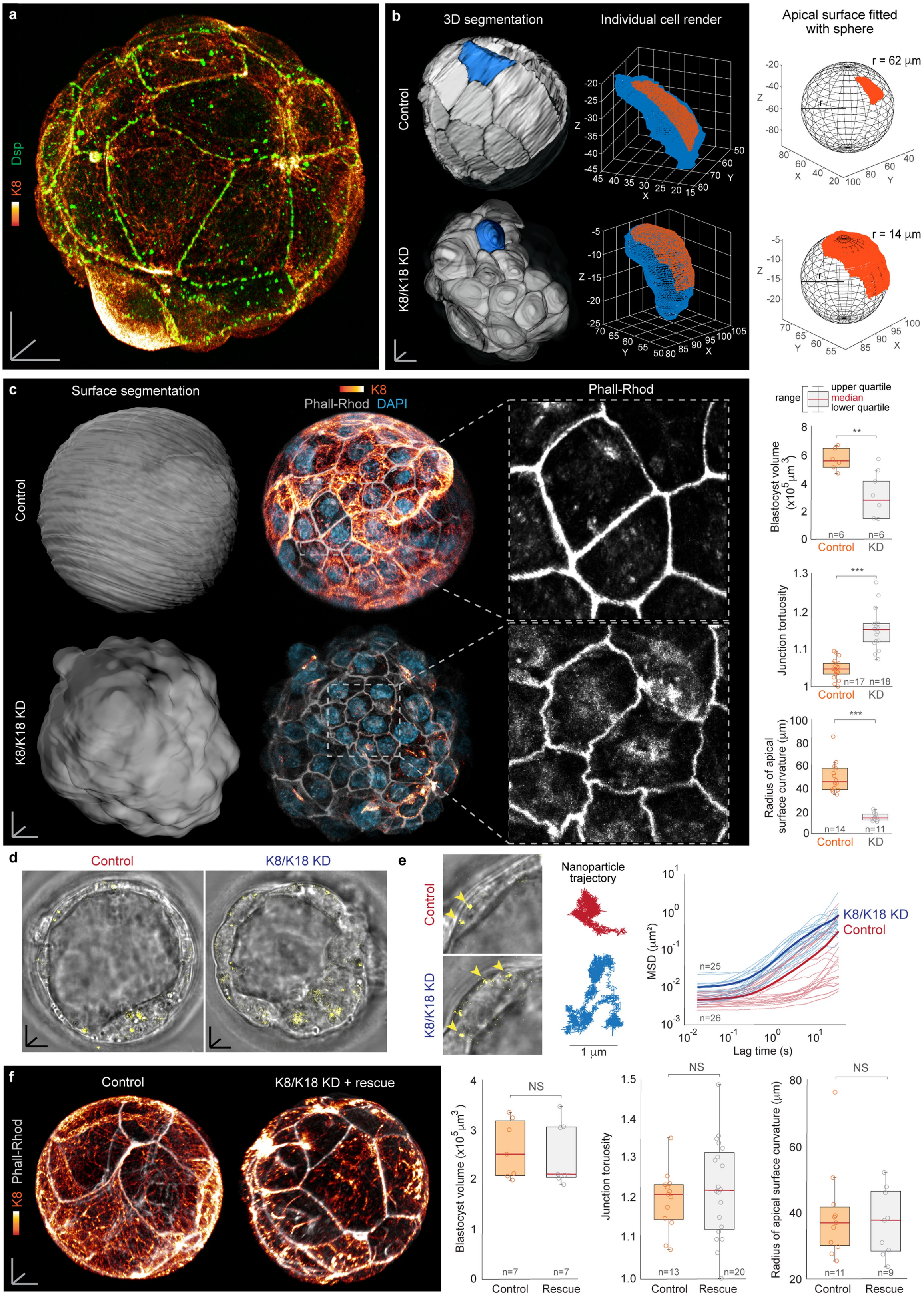
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Keratins promote actin stability and apical polarization. a, FRAP experiments for mRuby2-Actin performed at the apical domain of interphase cells with keratins, cells without keratins, and cells microinjected with desmosome siRNAs. Selected photobleached regions of interest (left) and kymographs of pre- and post-FRAP fluorescence intensities (right) are shown. Data are from three independent experiments. **b**, Analysis of FRAP experiments. Graphs show the fluorescence recovery of mRuby2-Actin over time for each condition. Thinner red lines indicate raw data after normalization, thicker red lines are fitted exponential curves, and thick black lines represent the mean fitted exponential curves. **c**, Cells lacking keratins and cells with reduced desmosome expression show a smaller immobile fraction of mRuby2-Actin compared to cells with keratins. $^{**}P = 0.0002$ for without keratins; $^{**}P = 0.003$ for desmosome KD; Kruskal–Wallis test. **d**, Immunofluorescence of 16-cell stage control embryos and embryos treated with cytochalasin D and SiR-Actin. Disruption of actin stability using cytochalasin D reduces accumulation of apical polarity markers PARD6B and PKCζ. By contrast, increasing actin stability using SiR-Actin increases apical polarity levels. $^{*}P = 0.03$; $^{***}P = 0.0009$; ANOVA test for PARD6B; $^{*}P = 0.03$; $^{**}P = 0.003$; Kruskal–Wallis test for PKCζ. **e**, Desmosome knockdown in 16-cell stage embryos reduces levels of apical polarity markers PARD6B and PKCζ. $^{***}P = 0.0002$ for PARD6B; $^{***}P = 0.001$ for PKCζ; Unpaired, two-tailed Mann–Whitney *U*-test. **f**, Live imaging of K18-Emerald in an embryo displaying a cell division. After division, the daughter cell that did not inherit keratins (cyan) undergoes apical constriction to form the pluripotent inner cell mass[35], whereas the outer daughter cell that inherited keratins (yellow) does not internalize. Data are from three independent experiments. **g**, Analysis of internalization events in cells that inherited (K+) or did not inherit (K−) keratin filaments after division. $^{***}P < 0.0001$; two-tailed Fisher's exact test. **h**, Immunofluorescence of endogenous K8 and AMOT in a 16-cell stage embryo. Right panels indicate zoomed views of the apical region of cells with and without keratins, with separate K8 and AMOT channels for better visualization. Cells with keratins display higher levels of apical AMOT than cells lacking keratins and cells with K8 and K18 knockdown. $^{*}P = 0.04$; $^{***}P < 0.0001$; Kruskal–Wallis test. Scale bars, 10 μm.



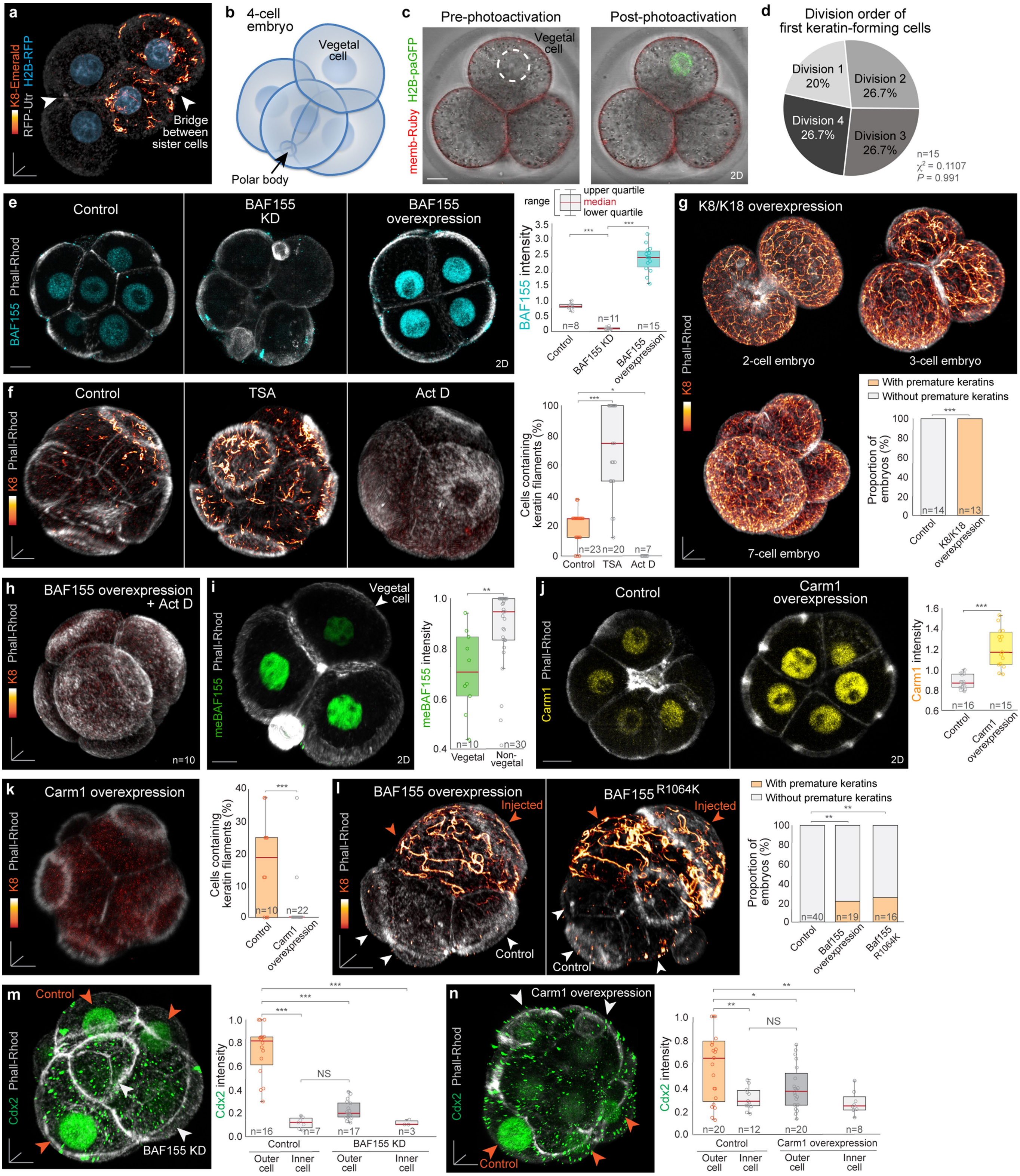
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Experimental manipulations of keratin levels show that keratins regulate CDX2 to specify the first trophectoderm cells of the embryo. a, Immunofluorescence for K8 in embryos microinjected with siRNAs for K8 and K18 at the one-cell stage, or into only one cell at the two-cell stage. This double-knockdown approach extensively eliminates keratin filament assembly. Data are from five independent experiments. **b**, Knockdown of K8 and K18 in half of the embryo also eliminates filament formation by K19. White arrowheads show knockdown cells. Data are from three independent experiments. **c**, Keratin overexpression causes a premature and widespread assembly of a keratin network within the 8- to 16-cell stage embryo. Images show examples of embryos microinjected with high levels of *K8* and *K18* RNA at the 1-cell stage, or into one cell of the 2-cell embryo. Data are from three independent experiments. **d**, Keratin overexpression causes some filaments to be inherited by inner cells of the 16-cell stage embryo (yellow segmented cell indicated by arrow in left panel). 2D view shows keratin filament organization within outer and inner cells of keratin overexpressing embryos (right). Data are from three independent experiments. **e**, Inner cells in keratin overexpressing embryos express lower levels of nuclear YAP and CDX2 than outer cells. $^{*}P = 0.04$; $^{***}P < 0.0001$; unpaired, two-tailed Mann–Whitney *U*-test. **f**, Knockdown of YAP using siRNAs microinjected into one cell of the two-cell embryo reduces CDX2 levels, in both keratin-positive and keratin-negative cells. H2B-RFP was co-injected with the siRNAs to identify the knockdown cells (white arrowheads). $^{***}P < 0.0001$; ANOVA test for CDX2. Right graph shows that our knockdown approach using *YAP* siRNAs effectively reduced YAP levels. $^{*}P = 0.03$, unpaired, two-tailed Mann–Whitney *U*-test for YAP. **g**, Scheme depicting cloning strategy to generate rescue constructs for K8 and K18. The coding regions of K8 and K18 are indicated by thick yellow arrows, and the targeted sequence locations for the keratin siRNAs used in this study are indicated by the red arrows. The specific siRNA target sequences are highlighted in yellow, corresponding to the keratin wild-type (WT) sequence (top rows). The rescue construct sequences are indicated (bottom rows). Note the conservation of amino acid sequence despite the scrambling of DNA bases throughout the siRNA target sequence. In each experiment, H2B-RFP was co-injected with the siRNAs and/or mRNAs to label the injected half of the embryo, and 100% of H2B-positive cells displayed keratin filaments when injected with the rescue construct. K8/K18-knockdown cells express lower levels of CDX2 than control cells with keratins, but this phenotype is rescued when the keratin rescue constructs are co-injected with keratin siRNAs. $^{***}P < 0.0001$; ANOVA test. Scale bars, 10 µm.



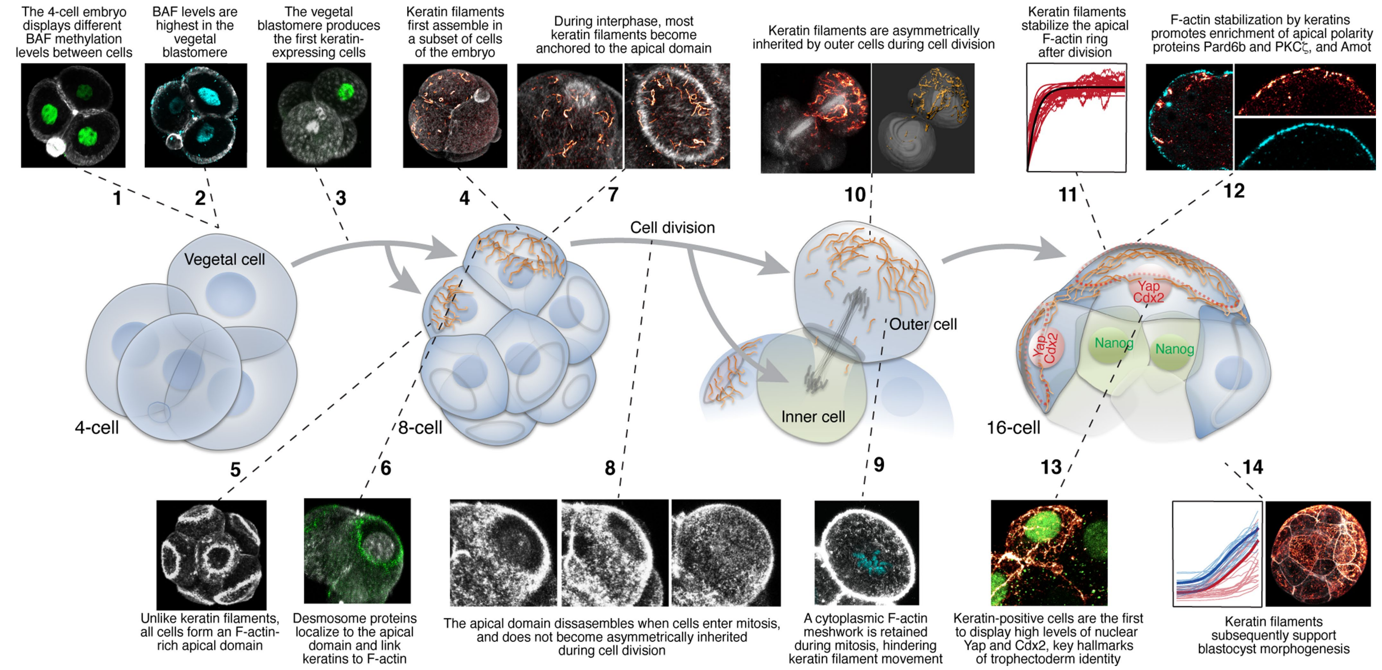
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Keratins regulate blastocyst morphogenesis. **a**, Punctate desmosome structures labelled using immunofluorescence for desmoplakin (Dsp) colocalize with K8 along the trophectoderm cell–cell junctions of the blastocyst. Data are from three independent experiments. **b**, Analysis of apical surface curvature in control and K8/K18-knockdown blastocysts. Individual cells within the intact embryo were computationally-segmented in 3D. Single cells (blue) are selected for apical surface analysis. Middle panels show rendering of the apical surfaces (orange) of the selected cells. The right panels show fitting of the cell apical surface to a sphere for calculation of radius of apical surface curvature. Data are from three independent experiments. **c**, K8/K18-knockdown blastocysts display morphogenetic defects, revealed by smaller blastocyst volume, higher junctional tortuosity, and trophectoderm cells with lower radius of apical surface curvature. $^{**}P = 0.004$ for blastocyst volume; $^{***}P < 0.0001$ for junctional tortuosity; $^{***}P < 0.0001$ for surface curvature; unpaired, two-tailed Mann–Whitney *U*-test. **d**, 2D confocal planes of live control and K8/K18-knockdown blastocysts, microinjected with fluorescent nanoparticles (yellow). Data are from three independent experiments. **e**, Images show nanoparticles within single trophectoderm cells, in control and K8/K18-knockdown embryos. Middle panels show representative trajectories of nanoparticle movement. Graph shows their mean squared displacement (MSD) over lag time. Thicker lines represent the mean of individual curves. The graph has two phases revealing different cytoskeletal properties: a time-independent (short lag times) and a time-dependent (long lag times) phase. These phases are associated with elasticity and viscosity, respectively[40]. Differences in MSD during the time-independent phase reveal higher elasticity, indicative of lower cytoplasmic stiffness, in the K8/K18-knockdown cells. **f**, Co-injection of keratin rescue constructs with K8/K18 siRNAs can restore blastocyst morphology to control conditions. Unpaired, two-tailed Student's *t*-test for blastocyst volume and surface curvature; unpaired, two-tailed Mann–Whitney *U*-test for junction tortuosity; NS, not significant. Scale bars, 10 µm.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Heterogeneities in BAF155 and CARM1 within the early embryo trigger differential expression of keratins at the eight-cell stage. a, Live-imaging of an embryo expressing K8-Emerald, H2B-RFP and RFP-Utrophin confirms that the first cells to assemble keratin filaments are sister cells. The microtubule bridge connecting sister cells can be identified by RFP-Utrophin accumulation (white arrowheads)[41]. Data are from three independent experiments. **b**, Scheme shows the stereotypical 3D organization of a tetrahedral four-cell embryo. The vegetal blastomere is located distal from the polar body. **c**, Selective photoactivation of the vegetal blastomere. The vegetal blastomere is identified based on its distal position from the polar body. The vegetal cell nucleus is then targeted with a two-photon laser (820 nm light) to photoactivate H2B-paGFP. 2D confocal planes show efficient photoactivation immediately after 820 nm light illumination. Data are from three independent experiments. **d**, The first cells to form keratin filaments are unrelated to the order of cell divisions during the 4- to 8-cell stage transition. χ^2 test. **e**, BAF155 knockdown reduces BAF155 immunofluorescence levels relative to control blastomeres, while BAF155 overexpression increases them. Embryos were microinjected with *BAF155* siRNAs or high levels of *BAF155* RNA respectively at the one-cell stage. $^{***}P < 0.0001$; ANOVA test. **f**, Embryos treated with trichostatin A (TSA) display extensive keratin filament formation, while embryos treated with actinomycin D (Act D) do not form filaments. $^{*}P = 0.0489$; $^{***}P < 0.0001$; ANOVA test. **g**, Microinjection of *K8* and *K18* mRNA into the one-cell embryo causes premature assembly of an extensive keratin filament network throughout early blastomeres before the eight-cell stage. $^{***}P < 0.0001$; two-sided Fisher's exact test. **h**, BAF155-overexpressing embryos treated with actinomycin D do not form keratin filaments at the eight-cell stage. Data are from three independent experiments. **i**, Dimethyl-BAF155 is lowest in the vegetal blastomere. $^{**}P = 0.004$; unpaired, two-tailed Student's *t*-test. **j**, CARM1 overexpression increases CARM1 immunofluorescence levels relative to control blastomeres. Embryos were microinjected with high levels of *Carm1* RNA at the 1-cell stage. $^{***}P < 0.0001$; unpaired, two-tailed Mann–Whitney *U*-test. **k**, CARM1 overexpression reduces keratin filament assembly. $^{***}P = 0.0007$; unpaired, two-tailed Student's *t*-test. **l**, Overexpression of BAF155 or mutant BAF155(R1064K) causes premature keratin filament assembly at the four-cell stage. $^{**}P = 0.009$ for BAF155 overexpression; $^{**}P = 0.005$ for BAF155(R1064K); two-sided Fisher's exact test. **m**, BAF155-knockdown blastomeres (white arrowheads) display lower levels of CDX2 than control cells (orange arrowheads) at the same stage. *BAF155* siRNAs were microinjected into only one cell of the two-cell embryo. $^{***}P < 0.0001$; ANOVA test. **n**, CARM1-overexpression blastomeres (white arrowheads) display lower levels of CDX2 than control blastomeres (orange arrowheads) at the same stage. High levels of *Carm1* RNA were microinjected into only one cell at the two-cell stage. $^{**}P = 0.005$ for control inner cells; $^{*}P = 0.04$ for CARM1 overexpression outer cells; $^{**}P = 0.006$ for CARM1-overexpression inner cells; ANOVA test. Scale bars, 10 μm.



Extended Data Fig. 10 | Scheme summarizing the main findings. Keratin expression is regulated by early heterogeneities in the BAF complex. During inner–outer cell segregation, apically localized keratin filaments are asymmetrically inherited by outer daughter cells, where they stabilize apical F-actin to promote apical polarity and acquisition of a trophectoderm fate. At late stages, keratins also support blastocyst morphogenesis. The numbers indicate the key events.

nature research

Corresponding author(s): Nicolas Plachta
Last updated by author(s): Jun 2, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Time-lapse imaging of live samples and single confocal scans of fixed immunostained samples were both performed using Zeiss ZEN software, on a Zeiss LSM 780 or LSM 880.
Data analysis	All image analyses were performed using Imaris 8.2 (Bitplane AG), Fiji, and MATLAB (Version R2018a). Statistical analyses were performed using GraphPad Prism (Version 8.3) and Microsoft Excel (Version 16.3). Custom code has been deposited in a publicly available repository at https://github.com/gracelhy/Analysis-of-embryo-parameters

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Source Data behind Figs. 1 to 4 and Extended Data Figs. 3 to 10 are available within the manuscript files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical test was performed to determine sample size. Sample size was determined based on prior experience and typical ranges used by research groups in the preimplantation mouse embryo field, and also in accordance to statistical test requirements. Previous work utilizing similar sample sizes includes Zenker et al. 2018 Cell 173:776-791; Zenker et al. 2017 Science 357: 925-928; White et al. 2016 Cell 165(1): 75-87.
Data exclusions	Embryos excluded from analyses include: 1) Unsuccessfully microinjected embryos that display low or undetectable fluorescence labeling unsuitable for quantitative analysis, and 2) 10-15% of embryos that display arrested or slower development in culture conditions. These exclusion criteria have been utilized in our previous work.
Replication	All experiments in this study were successfully performed at least 3 times with different batches of embryos, mRNA or siRNA preparations.
Randomization	Embryos were randomly allocated into experimental groups. All embryos and cells within embryos were randomly selected for analysis.
Blinding	Successfully developed and imaged embryos have to be selected for subsequent analysis. Therefore, investigators were only blinded for computational analysis following acquisition of imaging data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☐	☒	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	Primary antibodies: Rat monoclonal anti-Keratin 8 (DSHB, TROMA-I) used at 1:20. Rabbit polyclonal anti-Keratin 18 (Sigma, SAB4501665, Lot #310243) used at 1:200. Rat monoclonal anti-Keratin 19 (DSHB, TROMA-III) used at 1:50. Mouse monoclonal anti-Pan-keratin (Cell Signaling, 4545) used at 1:50. Mouse monoclonal anti-alpha-tubulin (Sigma, T6199) used at 1:1000. Mouse monoclonal anti-Pard6b (Santa Cruz, 166405) used at 1:50. Mouse monoclonal anti-PKCzeta (Santa Cruz, 17781) used at 1:50. Rabbit monoclonal anti-Yap (Cell Signaling, 8418S) used at 1:500. Rabbit polyclonal anti-Cdx2 (Abcam, 88129) used at 1:200. Rabbit polyclonal anti-Nanog (Abcam, 80892) used at 1:200. Guinea pig polyclonal anti-Desmoplakin1 (Progen, DP-1) used at 1:100. Mouse monoclonal anti-Desmoglein1/2 (Progen, 61002S) used undiluted. Mouse monoclonal anti-Plakophilin2 (Progen, 651101S) used undiluted. Mouse monoclonal anti-Plakoglobin (Progen, 61005S), used undiluted. Mouse monoclonal anti-BAF155 (Santa Cruz, 48350) used at 1:50. Rabbit monoclonal anti-Carm1 (Cell Signaling, 3379S) used at 1:150. Rabbit polyclonal anti-dimethyl-BAF155 Arg1064 (Merck, ABE1339, Lot #3174767) used at 1:100. Secondary antibodies:

	Alexa Fluor 488 Goat anti-Mouse (Invitrogen) used at 1:500. Alexa Fluor 488 Goat anti-Rat (Invitrogen) used at 1:500. Alexa Fluor 488 Goat anti-Rabbit (Invitrogen) used at 1:500. Alexa Fluor 647 Goat anti-Mouse (Invitrogen) used at 1:500. Alexa Fluor 647 Goat anti-Rat (Invitrogen) used at 1:500. Alexa Fluor 647 Goat anti-Rabbit (Invitrogen) used at 1:500. Alexa Fluor 488 Donkey anti-Guinea pig (Jackson ImmunoResearch, 706-545-148) used at 1:500.
Validation	All antibodies were previously validated by vendors and/or published work. Relevant studies include: Keratin 8: filamentous keratin network in ES cells (Schwarz et al. 2015 Sci. Rep. 5: 9007); keratin network in mouse blastocysts showing trophectoderm-specific localization (Ralston and Rossant 2008 Dev. Biol. 313: 614-629). Keratin 18: Validation by manufacturer shows specific filamentous staining of HeLa cells in the presence of the antibody; specific staining in notochordal cells (Rodrigues-Pinto et al. 2016 J Orthop Res 34(8):1327-1340). Keratin 19: Expression in mouse embryonic mammary glands, as expected and colocalizing with other keratin subtypes within the same tissue (Sun et al. 2010 Histochem Cell Biol 133:213-221). Pan-keratin: Expression in cervical carcinoma tissues, with keratins as known biomarkers (He et al. 2019 Cell Rep 26(10):2636-2650); Specific staining at the interface of a tumor-stromal in vitro assay (Begum et al. 2019 Sci Rep 9:11187). Alpha-tubulin: Specific localization to interphase and cytokinetic microtubule bridges in mouse embryos (Zenker et al. 2017 Science 357: 925-928); specific localization to mitotic spindles in mouse embryos (Zenker et al. 2018 Cell 173: 776-791). Pard6b: Specific apical localization in polarized cells, and disruption of cell organization results in its mislocalization (Choi et al. 2019 J Cell Biol 218(7): 2277-2293). PKCzeta: Staining shows specific apical localization in compacted 8-cell blastomeres of the mouse embryo (Zhu et al. 2017 Nat. Comms. 8:1-16); specific enrichment at apical membranes of outer cells of the mouse embryo, which is disrupted by Pard6b shRNA (Alarcon 2010 Biol. Reprod. 83:347-358). Yap: Specific staining of nuclei of polar outer cells of the mouse embryo (Anani et al. 2014 Development 141:2813–2824). Cdx2: Specific staining of nuclei in mouse embryos throughout preimplantation development, and elevated levels in trophectoderm cells of the blastocyst relative to those of the inner cell mass (White et al. 2016 Cell 165:75-87); specific staining of nuclei in morula-stage embryos (Samarage et al. 2015 Dev. Cell 34:435-447). Nanog: Specific staining of nuclei in mouse embryos (White et al. 2016 Cell 165:75-87); specific staining of inner cell mass nuclei in mouse blastocysts (Panamarova et al. 2016 Development 143:1271-1283). Desmoplakin1: Punctate staining along cell-cell junctions of trophectoderm cells in the blastocyst, as expected for desmosomal complexes in epithelia (Schwarz et al. 2015 Sci. Rep. 5:9007). Desmoglein1/2: Specific punctate staining along cell-cell junctions of keratinocytes, as expected (Price et al. 2018 Nat Comms 9:5284). Plakophilin2: Specific localization along cell-cell junctions of cardiomyocytes (Merkel et al. 2019 Mol Biol Cell 30(21):2639-2650). Plakoglobin: Membrane localization along cell-cell junctions of keratinocytes, as expected (Dayal et al. 2014 J Cell Sci 127:740-751). BAF155: Validation by manufacturer shows specific nuclei staining in fixed HeLa cells. Carm1: Specific staining of nuclei in mouse embryos (White et al. 2016 Cell 165:75-87). Dimethyl-BAF155: Staining of normal and tumor breast tissue sections (Wang et al. 2014 Cancer Cell 25:21-36).

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	C57BL/6 wild-type female mice were superovulated using 5 iu of pregnant mare serum (PMS, National Hormone and Peptide Program) gonadotropin given intraperitoneally and 5 iu of recombinant chorionic gonadotrophin (CG, National Hormone and Peptide Program) given 48 h after and immediately before mating, according to animal ethics guidelines of the Agency for Science, Technology and Research, Singapore.
Wild animals	No wild animals were used.
Field-collected samples	No field-collected samples were used.
Ethics oversight	Mouse embryo work was performed according to animal ethics guidelines of the Agency for Science, Technology and Research, Singapore. All protocols (IACUC #181370) were approved by the Biological Resource Centre IACUC Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Embryo donors were carefully selected to meet strict research inclusion criteria to minimize potential risks to the donors. Female donors were between the ages of 20 to 40 years old and had at least one healthy baby at the Center for Reproductive Medicine, Sixth Affiliated Hospital of Sun Yat-Sen University.
Recruitment	Participation in this study was entirely voluntary and no financial inducements were given for embryo donation. Embryo donors were informed that their donated surplus embryos would be used to study the developmental mechanisms of human embryos and that their donation would not affect their IVF cycle. All embryo donors signed informed consent forms stating clearly the goals of the research, potential benefits and risks, and steps taken to ensure that their privacy was met. Conditions of donation: not for profit. Given that only surplus embryos that were not used for IVF could be donated to this study, there could be a bias in

	terms of the quality or condition of embryos used for research, and it is unclear whether these surplus embryos could develop to term if implanted in the uterus.
Ethics oversight	This work was approved by the Ethics Committee of Center for Reproductive Medicine, Sixth Affiliated Hospital of Sun Yat-Sen University (Research license 2019SZZX-008). The Medicine Ethics Committee of Center for Reproductive Medicine, Sixth Affiliated Hospital of Sun Yat-Sen University is composed of 11 members, including experts of laws, scientists and clinicians with relevant expertise. The Committee evaluated the scientific merit and ethical justification of this study and conducted a full review of the donations and use of these samples. All informed consent and research procedures were carried out in accordance to the ethical and regulatory framework set forth by the Ministry of Science and Technology and the Ministry of Health of the People's Republic of China. They have also been approved by the Ethics Committee of the Reproductive Medicine and Prenatal Diagnosis of the 6th Affiliated Hospital of Sun Yat-sen University, and the Ethics Committee of Institute of Zoology, Chinese Academy of Sciences.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Effect of non-pharmaceutical interventions to contain COVID-19 in China

https://doi.org/10.1038/s41586-020-2293-x

Received: 12 March 2020

Accepted: 23 April 2020

Shengjie Lai[1,2,10]✉, Nick W. Ruktanonchai[1,3,10]✉, Liangcai Zhou[4], Olivia Prosper[5], Wei Luo[6,7], Jessica R. Floyd[1], Amy Wesolowski[8], Mauricio Santillana[6,7], Chi Zhang[9], Xiangjun Du[9], Hongjie Yu[2] & Andrew J. Tatem[1]✉

On 11 March 2020, the World Health Organization (WHO) declared coronavirus disease 2019 (COVID-19) a pandemic[1]. The strategies based on non-pharmaceutical interventions that were used to contain the outbreak in China appear to be effective[2], but quantitative research is still needed to assess the efficacy of non-pharmaceutical interventions and their timings[3]. Here, using epidemiological data on COVID-19 and anonymized data on human movement[4,5], we develop a modelling framework that uses daily travel networks to simulate different outbreak and intervention scenarios across China. We estimate that there were a total of 114,325 cases of COVID-19 (interquartile range 76,776–164,576) in mainland China as of 29 February 2020. Without non-pharmaceutical interventions, we predict that the number of cases would have been 67-fold higher (interquartile range 44–94-fold) by 29 February 2020, and we find that the effectiveness of different interventions varied. We estimate that early detection and isolation of cases prevented more infections than did travel restrictions and contact reductions, but that a combination of non-pharmaceutical interventions achieved the strongest and most rapid effect. According to our model, the lifting of travel restrictions from 17 February 2020 does not lead to an increase in cases across China if social distancing interventions can be maintained, even at a limited level of an on average 25% reduction in contact between individuals that continues until late April. These findings improve our understanding of the effects of non-pharmaceutical interventions on COVID-19, and will inform response efforts across the world.

As of 30 March 2020, the outbreak of COVID-19, which is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has resulted in 693,282 confirmed cases and 33,106 deaths across the world[6]. As the disease has only recently emerged, effective pharmaceutical interventions are not expected to be available for months[7], and healthcare resources will be limited for treating all cases. Non-pharmaceutical interventions (NPIs) are therefore essential components of the public health response to COVID-19 outbreaks[6,8–10]. These include the isolation of individuals who are ill, contact tracing, quarantine of exposed individuals, travel restrictions, school and workplace closures, cancellation of mass gatherings, and hand-washing, among others[8–10]. Such measures aim to reduce the transmission of the virus by delaying the timing and reducing the size of the peak of the epidemic, thus buying time for preparations to be made in the healthcare system and creating the potential for vaccines and drugs to be used at a later stage[8].

Three major groups of NPIs have been implemented to contain the spread and reduce the size of the outbreak of COVID-19 across China[11]. First, intercity travel restrictions were used to prevent further seeding of the virus during the Chinese New Year holiday period. A cordon sanitaire of Wuhan and surrounding cities in Hubei province was put in place on 23 January 2020, two days before the Chinese New Year, which started on 25 January 2020. After this date, travel restrictions were also put in place in other provinces across the country. Second, the early identification and isolation of cases was prioritized, including improving the screening, identification, diagnosis, isolation, reporting and contact tracing of people who were suspected or confirmed to have the disease[11]. Local governments across China encouraged and supported the routine screening and quarantine of travellers from Hubei province in an attempt to detect COVID-19 infections as early as possible. The average interval from the onset of symptoms to laboratory confirmation dropped from 12 days in the early stages of the outbreak to 3 days in early February, indicating that these efforts improved detection and diagnosis[3,12]. Third, contact restrictions and social distancing measures, together with personal preventive actions such as hand-washing, were implemented to reduce the risk of exposure at the community level. As part of these social distancing policies, the Chinese government encouraged people to stay at home as much as possible, cancelled or postponed large public events and mass gatherings, and closed libraries, museums and workplaces[13,14]. School holidays were also extended, with the end date of the Chinese New Year holiday period changed from 30 January 2020 to 10 March 2020 for Hubei province, and to 9 February 2020 for many other provinces[15,16].

---

[1]WorldPop, School of Geography and Environmental Science, University of Southampton, Southampton, UK. [2]School of Public Health, Fudan University, Key Laboratory of Public Health Safety, Ministry of Education, Shanghai, China. [3]Population Health Sciences, Virginia Tech, Blacksburg, VA, USA. [4]Wuhan Center for Disease Control and Prevention, Wuhan, China. [5]Department of Mathematics, University of Tennessee, Knoxville, TN, USA. [6]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA. [7]Department of Pediatrics, Harvard Medical School, Boston, MA, USA. [8]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [9]School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, China. [10]These authors contributed equally: Shengjie Lai, Nick W. Ruktanonchai. ✉e-mail: shengjie.lai@soton.ac.uk; nr1e14@soton.ac.uk; a.j.tatem@soton.ac.uk

The implementation of these NPIs coincided with a rapid decline in the number of new cases across China, albeit at high economic and social costs[3,12]. Previous studies have examined the effects of the lockdown of Wuhan[17,18], travel restrictions[19], airport screening[20], isolation of cases and contact tracing on the containment of the disease[21]. However, a comprehensive and quantitative comparison of the effectiveness of different NPIs, and the time at which they were implemented, for containing the outbreak of COVID-19 in China is lacking. On the basis of epidemiological data on COVID-19 and historical and near-real-time anonymized data on human movement, we developed a stochastic susceptible–exposed–infectious–removed (SEIR) modelling framework based on travel networks to simulate the spread of COVID-19 across 340 prefecture-level cities in mainland China. Within each city, we estimated the numbers of susceptible, exposed, infectious, and recovered/removed ('removed' refers to the individuals who were isolated to prevent further transmission, and deceased individuals) people per day from 1 December 2019. Using this modelling framework, we conducted before-and-after comparable analyses to quantify the relative effect of the three major groups of NPIs—that is, the restriction of intercity population movement, the identification and isolation of cases, and the reduction of travel and contact within cities to increase social distance—in China. We also assessed the risk of COVID-19 transmission since the lifting of travel restrictions on 17 February 2020.

Reconstructing the spread of COVID-19

The epidemiological parameters that were estimated for the early stage of the outbreak in Wuhan were initially used to parameterize the epidemic before interventions were widely implemented[5]. The three major groups of NPIs outlined above were derived and measured using data on population movement between and within cities (obtained from smartphone users of Baidu location-based services[4]) and data on the delay between the onset of illness and the reporting of cases across the country. Population travel and contact patterns changed substantially after the implementation of interventions, and the timeliness of case reporting also improved (Fig. 1, Supplementary Tables 1, 2). These indicators were then incorporated into the model (see Methods).

We estimated that there were a total of 114,325 cases of COVID-19 (interquartile range (IQR) 76,776–164,576) in mainland China as of 29 February 2020, 85% of which were in Hubei province (Extended Data Table 1). The outbreak increased exponentially before Chinese New Year, but the peaks of epidemics across the country quickly appeared around the time of Chinese New Year after the implementation of NPIs. The estimated epidemics and peaks were consistent with patterns of reported data by onset date, with strong correlations between daily estimates and reported data across time and regions (Extended Data Fig. 1). The overall correlation between the estimated number of cases and the reported number by province, as of 29 February 2020, was also significant ($P < 0.001$, $R^2 = 0.86$), with a high sensitivity (91%, 280/308) and specificity (69%, 22/32) in predicting cities with or without cases of COVID-19 (Extended Data Fig. 1a, b).

Quantifying the effect of different NPIs

Without NPIs, our model predicted the number of cases of COVID-19 to increase rapidly across China, with a 51-fold (IQR 33–71) increase in Wuhan, a 92-fold (58–133) increase in other cities in Hubei province and a 125-fold (77–180) increase in other provinces by 29 February 2020. However, the apparent effectiveness of different interventions varied (Fig. 2). The lockdown of Wuhan might not have prevented the seeding of the virus from the city, as the travel ban was put in place at the latter stages of population movement out of the city before Chinese New Year[22] (Fig. 1b). Nevertheless, if intercity travel restrictions had not been implemented, cities and provinces outside of Wuhan would have received more cases from Wuhan, and the affected geographical range would have expanded to the remote western areas of China (Extended Data Fig. 2c). In general, we estimated that the early detection and isolation of cases quickly and substantially prevented more infections than did the introduction of contact reduction and social distancing measures across the country (5-fold versus 2.6-fold). However, without the contact reduction intervention, in the longer term the epidemics would have increased exponentially across regions (Fig. 2c, f). Therefore, combined NPIs would bring about the strongest and most rapid effect on containment of the COVID-19 outbreak, with an interval of about one week between the introduction of NPIs and the peak of the epidemic (Extended Data Table 1).

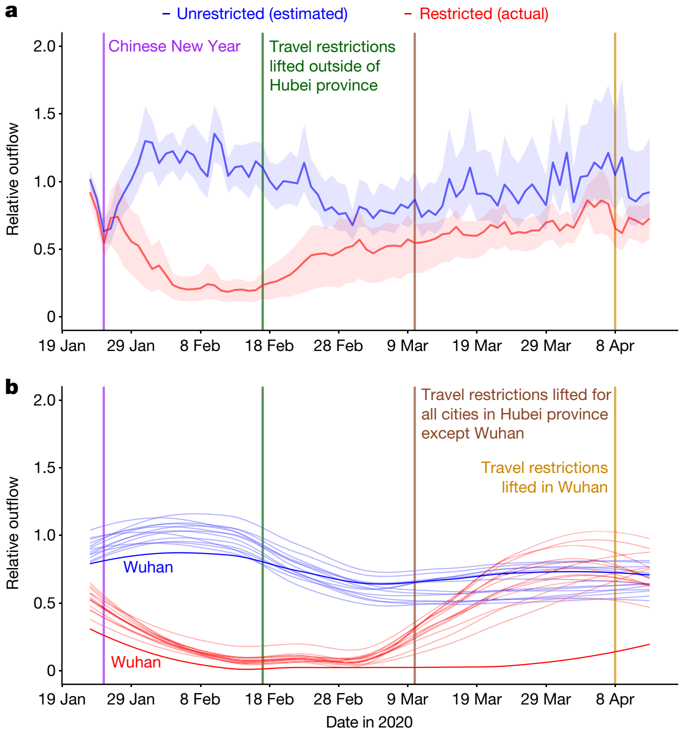


Fig. 1 | Relative daily volume of outbound travellers from cities across mainland China between 23 January and 13 April 2020. a, Relative outbound flows of travellers for all cities at prefecture level ($n = 340$) in mainland China, presented as the median (solid line) and IQR (shading). **b**, Relative outbound flows of travellers for cities in Hubei province. Wuhan is highlighted using darker colours. Each red line represents the outflow for each city in 2020, standardized by the mean of daily outflows for each city from 20 to 22 January 2020. Each blue line represents an estimate of the normal outflow by city under the scenario of no travel restrictions (on the basis of travel in previous years). The lines in **b** were smoothed using locally estimated scatterplot smoothing (LOESS) regression.

Timing of interventions

Our model suggests that, theoretically, if interventions in China had been implemented one week, two weeks or three weeks earlier than they actually were, the number of cases of COVID-19 could have been reduced by 66% (IQR 50–82%), 86% (81–90%) or 95% (93–97%), respectively (Fig. 3a). The geographical range of affected areas would also shrink from 308 cities to 192, 130 or 61 cities, respectively (Extended Data Fig. 3). However, if NPIs had been introduced one week, two weeks or three weeks later than they were, the number of cases might have increased by 3-fold (IQR 2–4), 7-fold (5–10) or 18-fold (11–26), respectively (Fig. 3b).

Lifting of travel restrictions

Under the interventions that were implemented from 17 February 2020—that is, the lifting of travel restrictions—the epidemics outside

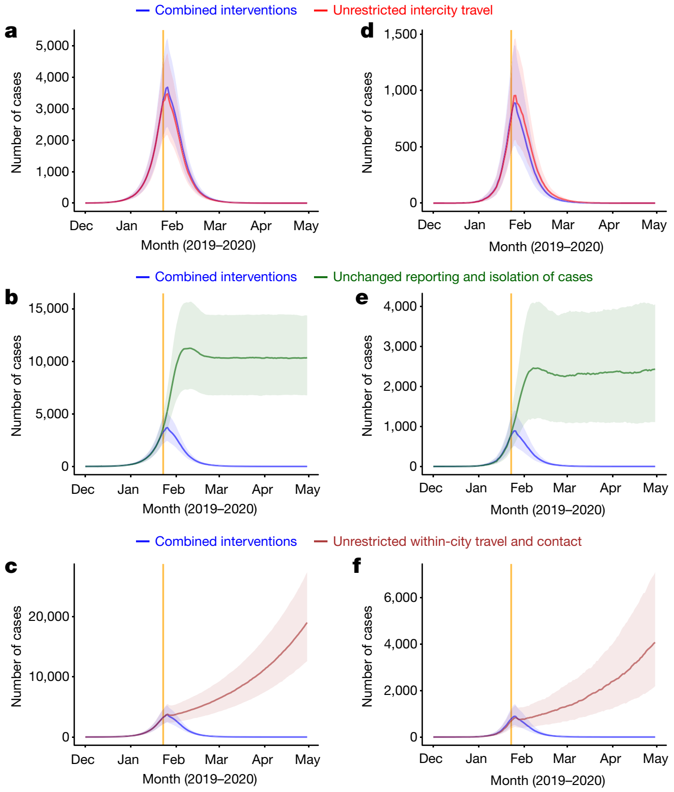


Fig. 2 | Estimated epidemic curves of the COVID-19 outbreak under various scenarios with or without NPIs by region. a–**c**, Estimates for the city of Wuhan. **d**–**f**, Estimates for cities outside of Hubei province in mainland China. The blue lines represent estimated transmission under combined NPIs, and the other coloured lines represent the scenario without one type of intervention. Data are presented as the median (solid line) and IQR (shading) of estimates (1,000 simulations). The orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020).

of Hubei province probably reached a low level (fewer than 10 cases per day, excluding imported cases from other countries) in early March, whereas Hubei province might need another four weeks to reach the same level as other provinces. However, if population contact resumed to normal levels, the lifting of travel restrictions might cause case numbers to rise again (Fig. 3c). Accordingly, our simulations suggest that maintaining social distancing even to a limited extent (for example, a 25% reduction in contact between individuals on average) through to late April would help to ensure control of COVID-19 in epicentres such as Wuhan.

Our estimates were sensitive to the basic reproduction number (R_0); under a higher R_0 value, the peaks of epidemics were higher and later, and more time was needed to contain the outbreak (Extended Data Fig. 3). Sensitivity analyses also suggested that our model could have robustly measured relative changes in the efficacy of interventions under different epidemiological parameters and transmission scenarios (Extended Data Figs. 4–9).

Discussion

Our findings show that combined NPIs substantially reduced the transmission of COVID-19 across China. Earlier implementation of NPIs could have notably reduced the magnitude and geographical range of the outbreak, but—equally—a delayed response would have led to a larger outbreak. China's aggressive, multifaceted response is likely to have prevented a far worse situation, which would have accelerated the spread of the virus globally. The evidence from China provides information that will be of use in efforts to contain the spread of COVID-19 and mitigate the effects of the disease in other regions around the world[3,12].

Our results suggest three key points. First, they support and validate the idea that population movement and close contact has a major role in the spread of COVID-19 within and beyond China[22,23]. As the lockdown of Wuhan happened at the latter stages of population movement before Chinese New Year, travel restrictions did not halt the seeding of the virus from Wuhan, but did prevent cases being exported from Wuhan to a wider area. Second, the importance and effects of the three types of NPIs differed. Compared with travel restrictions, improved detection and isolation of cases, as well as social distancing, probably had a greater effect on the containment of the outbreak. The social distancing intervention reduced contact between people who travelled from the epicentre of the outbreak and other individuals. This is likely to have been especially helpful in curbing the spread of an emerging pathogen to the wider community, and to have reduced the risk of spread from asymptomatic or mild infections[8]. Third, given that travel and work have begun to resume in China, the country should consider at least the partial continuation of NPIs to ensure that the COVID-19 outbreak is sustainably controlled for the first wave of this outbreak. For example, the early identification and isolation of cases should be maintained—which might also help to prevent and delay the arrival of a second wave, considering the increasing numbers of cases that are imported from other countries and the presence of asymptomatic or subclinical infections in China[24].

The analyses presented here provide a comprehensive quantitative assessment of the effect of NPIs on the transmission of COVID-19. The model framework accounts for daily interactions of populations and interventions between and within cities, as well as the inherent

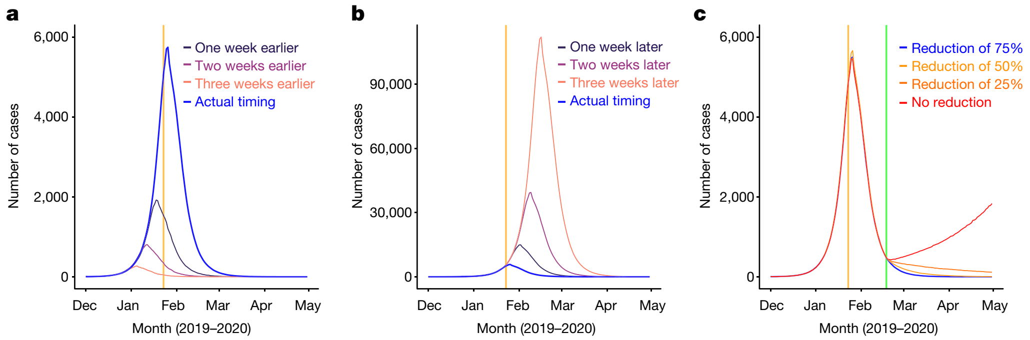


Fig. 3 | Estimates of the COVID-19 outbreak under various scenarios of intervention timing and lifting of travel restrictions across China. a, Estimated epidemic curves for interventions implemented earlier than their actual timing. **b**, Estimated epidemic curves for interventions implemented later than their actual timing. **c**, Estimated spread of COVID-19 for different rates of population contact after the lifting of intercity travel restrictions. The orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020), and the green line shows the date on which travel restrictions were lifted (17 February 2020).

statistical uncertainty that is associated with a paucity of epidemiological parameters before and after the implementation of interventions. The network-based SEIR model is methodologically robust and is built on the basic SEIR models that have been used previously to predict the transmission of COVID-19 in its early stages[23]. Considering the delays that exist in the reporting of cases, our approach can be used to enable a rapid, ongoing estimation of the effectiveness of various NPIs in different countries, and to aid decision-making relating to the control of outbreaks of COVID-19.

Our study has several limitations. First, our simulations were based on parameters that were estimated for symptomatic cases identified in the early stage of the outbreak in Wuhan, and might not account for asymptomatic and mild infections; we may therefore have underestimated the total number of infections. Second, our findings could be confounded by other factors that changed during the outbreak. Although we have shown that the apparent fall in the incidence of COVID-19 after Chinese New Year (25 January 2020) in China is likely to be attributed to the interventions taken, we cannot rule out the possibility that the decrease was partially attributable to other unknown seasonal factors—for example, temperature and absolute humidity[25,26]. Third, if the epidemiological parameters of COVID-19 transmission in other cities across China differed from the estimates[5]—which were based on the data in the early stage of the outbreak, when no NPIs were in place in Wuhan—then our estimates of the effectiveness of interventions in reducing the transmission of COVID-19 could be biased. Fourth, there are probably biases in population coverage, given that our model relies on data from mobile phone and Baidu users. Although a high percentage (from 46.9% in 2013 to 55.3% in 2018) of the population of China owns smartphones[27] (https://en.wikipedia.org/wiki/List_of_countries_by_smartphone_penetration), the group of mobile-phone users does not include specific subgroups of the population, particularly children. Therefore, our data on population movement may provide an incomplete picture, and differences between the characteristics of smartphone owners and non-owners may also bias our estimates. In addition, the magnitude and patterns of population movements could change year by year—although previous studies have suggested that travel patterns are consistent in their seasonality across years in China and other countries[22]. Finally, we only examined three main groups of NPIs, and other interventions might also have contributed to the containment of the outbreak. For example, owing to the sources of data that were available, we did not assess the effect of personal hygiene and protective equipment on containing the spread of COVID-19. Other sources of data and further investigations are needed to measure and evaluate the efficacy of each intervention.

COVID-19 has placed a substantial burden on health systems and society across many countries. From a public health standpoint, our results highlight that countries should consider proactively planning NPIs and relevant strategies for containment and mitigation, as the earlier implementation of NPIs could have led to substantial reductions in the size of the outbreak in China. Our results also provide guidance for countries as to the likely effectiveness of different NPIs at different stages of an outbreak. Suspected and confirmed cases of the disease should be identified, diagnosed, isolated and reported as early as possible to control the source of infection, and the implementation of cordon sanitaires or travel restrictions for areas that are heavily affected might prevent the virus spreading to wider regions. Reducing contact and increasing social distance between individuals, together with improved personal hygiene, can help to protect vulnerable populations and mitigate the spread of COVID-19 at the community level, and these interventions should be promoted throughout the outbreak to avoid resurgence. Our findings suggest that—as advocated by WHO—strategies that involve the early implementation of integrated NPIs should be prepared, deployed and adjusted to maximize the benefits of these interventions and minimize the health, social and economic effects of COVID-19 around the world[3].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2293-x.

1. World Health Organization. *WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020* https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (WHO, 2020).
2. Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China [in Chinese with English abstract]. *Zhonghua Liu Xing Bing Xue Za Zhi* **41**, 145–151 (2020).
3. World Health Organization. *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)* https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf (WHO, 2020).
4. *Baidu Migration* [in Chinese] https://qianxi.baidu.com/ (2020).
5. Li, Q. et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med*. **382**, 1199–1207 (2020).
6. World Health Organization. *Coronavirus disease (COVID-19) pandemic* https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (WHO, 2020).
7. Heymann, D. L. & Shindo, N. COVID-19: what is next for public health? *Lancet* **395**, 542–545 (2020).
8. Fong, M. W. et al. Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—social distancing measures. *Emerg. Infect. Dis*. **26**, 976–984 (2020).
9. Ryu, S. et al. Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—international travel-related measures. *Emerg. Infect. Dis*. **26**, 961–966 (2020).
10. Xiao, J. et al. Nonpharmaceutical measures for pandemic influenza in nonhealthcare settings—personal protective and environmental measures. *Emerg. Infect. Dis*. **26**, 967–975 (2020).
11. Chen, W. et al. Early containment strategies and core measures for prevention and control of novel coronavirus pneumonia in China [in Chinese with English abstract]. *Zhonghua Yu Fang Yi Xue Za Zhi* **54**, 239–244 (2020).
12. World Health Organization. *Press Conference of WHO-China Joint Mission on COVID-19* https://www.who.int/docs/default-source/coronaviruse/transcripts/joint-mission-press-conference-script-english-final.pdf?sfvrsn=51c90b9e_2 (2020).
13. The State Council of the People's Republic of China. *The announcement from Wuhan's headquarter on the novel coronavirus prevention and control* [in Chinese] http://www.gov.cn/xinwen/2020–01/23/content_5471751.htm (2020).
14. The State Council of the People's Republic of China. *The announcement on strengthening community prevention and control of pneumonia epidemic situation of new coronavirus infection* [in Chinese] http://www.gov.cn/zhengce/2020-01/27/content_5472516.htm (2020).
15. The State Council of the People's Republic of China. *The State Council's announcement on extending the 2020 Spring Festival Holiday* [in Chinese] http://www.gov.cn/zhengce/content/2020-01/27/content_5472352.htm (2020).
16. The People's Government of Shanghai Municipality. *The announcement on postponing the reoperation date of companies and the reopening date of schools* [in Chinese] http://www.shanghai.gov.cn/nw2/nw2314/nw2315/nw43978/u21aw1423601.html (2020).
17. Li, X., Zhao, X. & Sun, Y. The lockdown of Hubei Province causing different transmission dynamics of the novel coronavirus (2019-nCoV) in Wuhan and Beijing. Preprint at *medRxiv* https://doi.org/10.1101/2020.02.09.20021477 (2020).
18. Kraemer, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
19. Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
20. Quilty, B. J., Clifford, S., CMMID nCoV working group 2, Flasche, S. & Eggo, R. M. Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). *Euro Surveill*. **25**, 2000080 (2020).
21. Hellewell, J. et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**, e488–e496 (2020).
22. Lai, S. et al. Assessing spread risk of Wuhan novel coronavirus within and beyond China, January–April 2020: a travel network-based modelling study. Preprint at *medRxiv* https://doi.org/10.1101/2020.02.04.20020479 (2020).
23. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697 (2020).
24. National Health Commission of the People's Repulic of China. *Updates on pneumonia of new coronavirus infections as of March 31, 2020* [in Chinese] http://www.nhc.gov.cn/xcs/yqtb/202004/28668f987f3a4e58b1a2a75db60d8cf2.shtml (2020).
25. Wang, M. et al. Temperature significant change COVID-19 transmission in 429 cities. Preprint at *medRxiv* https://doi.org/10.1101/2020.02.22.20025791 (2020).
26. Luo, W. et al. The role of absolute humidity on transmission rates of the COVID-19 outbreak. Preprint at *medRxiv* https://doi.org/10.1101/2020.02.12.20022467 (2020).
27. Lai, S., Farnham, A., Ruktanonchai, N. W. & Tatem, A. J. Measuring mobility, disease connectivity and individual risk: a review of using mobile phone data and mHealth for travel medicine. *J. Travel Med*. **26**, taz019 (2019).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Model summary

An SEIR model based on travel networks was built to simulate the spread of COVID-19 between and within all prefecture-level cities in mainland China. This model has been made openly available for further use at https://github.com/wpgp/BEARmod. Population movement data across the country were used to estimate the intensity of travel restrictions and contact reductions. Data from illness onset to reporting of the first index case for each county were used to infer the changing timeliness of case identification and isolation across the course of the outbreak. The outputs of the model under NPIs were validated by using daily numbers of new cases reported across all regions in mainland China. On the basis of this modelling framework, the efficacy of applying or lifting non-pharmaceutical measures under various scenarios and timings were tested and quantified.

Data sources

Three datasets on population movement, which were obtained from Baidu location-based services that provide over 7 billion positioning requests per day[4,28], were used in this study to measure travel restrictions and social distancing across time and space. The first is an aggregated and de-identified dataset on near-real-time daily relative outbound and inbound flow of smartphone users for each prefecture-level city in 2020 (340 cities in mainland China were included) to understand patterns of mobility during the outbreak. The daily outflow from each city since the lockdown of Wuhan and the travel restrictions that were applied on 23 January 2020 were rescaled by the mean daily flow for each city from 20 to 22 January 2020 for comparing travel reductions across cities and years (Fig. 1).

The second Baidu dataset is a historical relative movement matrix with daily total number of users at the city level from 26 December 2014 to 26 May 2015, aligning with the 2020 Chinese New Year holiday period, for which the corresponding period is 1 December 2019 to 30 April 2020. We assumed that the pattern of population movements was the same in years when there were no outbreaks and interventions. Adjusted by the level of travel reductions derived from the 2020 dataset where applicable, the second dataset was used to simulate the spread of COVID-19 and predict transmission via population movements under various scenarios, with or without intercity travel restrictions. Corresponding city-level population data in 2015 for modelling were obtained from the Chinese Bureau of Statistics[29].

The third Baidu dataset measures daily population movements at the county level (2,862 counties in China) from 26 January to 30 April 2014, as described elsewhere[30]. On the basis of the assumption that the pattern of population contact was consistent across years when there were no interventions, it was used to estimate within-city travel and contact reduction during the outbreak and interventions. First, we aggregated data from county to city level and rescaled the daily flows from 29 January 2014 by the mean of the daily flow for 26–28 January period, aligning with the date of Wuhan's lockdown and the 2020 Chinese New Year holiday period. Then, the rescaled first dataset for 2020 under interventions was compared with the 2014 dataset to derive the percentage of travel decline for each city. The percentages for cities were averaged by day to preliminarily quantify the intensity of contact reduction in China under NPIs (Supplementary Table 2), as the policies of travel restriction and social distancing measures were implemented and occurred at the same time across the country.

We also collated data of the first case reported by county across mainland China to measure the delay from illness to case report as a reference of the improved timeliness of case identification, isolation and reporting during the outbreak (Supplementary Table 1). The daily number of COVID-19 cases by date of illness onset in the city of Wuhan, Hubei province and other provinces as of 13 February 2020 were used to further validate the epidemic curves estimated in this study across time. There was an abnormal increase of cases in Wuhan and Hubei province on 1 February 2020, on the basis of the date of illness onset[2]. We interpolated the number on 1 February 2020 by using the mean of numbers of cases reported on 31 January and 2 February 2020 in the epidemic curve. The number of cases reported by city across mainland China as of 29 February 2020 was used to define the predictability of our model across space. These case data were collated from the websites of national and local health authorities, news media and publications[2,3,31] (Supplementary Information).

Data analysis

We constructed a travel-network-based SEIR modelling framework (BEARmod) for before-and-after comparable analyses on the efficacy of NPIs. This model was extended from a typical SEIR model to specifically incorporate movement between locations that varied with each time step. In this model, each city was represented in the model as a separate subpopulation, with its own susceptible (S), exposed (E), infected (I) and recovered/removed (R) populations.

Exposure, infection and recovery

During each time step, infected people first recovered or were removed at an average rate r, where r was equal to the inverse of the average infectious period, and removal represents self-isolation and effective removal from the population as a potential transmitter of disease. This was incorporated as a Bernoulli trial for each infected person with a probability of recovering of $1 - \exp(-r)$. We used the median of time lags from illness onset to reported case as a proxy of the average infectious period, indicating the improving identification and isolation of cases under improved interventions (Supplementary Table 1). Then, the model converted exposed people to infectious by similarly incorporating a Bernoulli trial for each exposed individual, where the daily probability of becoming infectious $1 - \exp(-\varepsilon)$, where ε was the inverse of the average time spent exposed but not infectious, on the basis of the estimated incubation period (5.2 days, 95% confidence interval (CI) 4.1–7.0)[5]. Finally, to end the exposure, infection and recovery step of the model, the number of newly exposed people was calculated for each city on the basis of the number of infectious people in the city (I_i) and the average number of daily contacts that lead to transmission that each infectious person has (c). We simulated the number of exposed individuals in a patch on a given day through a random draw from a Poisson distribution for each infectious person, in which the mean number of new infections per person was c, which was then multiplied by the fraction of people in the city that were susceptible. We calculated the daily contact rate c using the basic reproduction rate that has been calculated in other studies ($R_0 = 2.2$ (95% CI 1.4–3.9)) divided by the average days (5.8, 95% CI 4.3–7.5) from onset to first medical visit and isolation[5], weighted by the relative level of daily contact where relevant, based on the Baidu movement data (Supplementary Table 2). Because simulation runs were not extended beyond five months, we did not include the addition of new susceptible people, or the conversion of recovered people back to susceptible.

The infection processes within each patch therefore approximate the following deterministic, continuous-time model, where c and r varied through time:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = S - c\frac{SI}{N}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = c\frac{SI}{N} - \varepsilon E$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \varepsilon E - rI$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = rI$$

Movement

After the model completed the infection-related processes, we moved infectious people between cities. To do this, we moved infected people from their current location to each possible destination (including remaining in the same place) using Bernoulli trials for each infected person, and each possible destination city. We parameterized the probability of moving from city i to city j (p_{ij}), which was equal to the proportion of smartphone users who went from city i to city j in the corresponding day from the Baidu dataset in 2015, accounting for the travel restrictions in 2020. This included modelling the numbers of people who stayed in the same location using p_{ii}, the proportion of users who did not move to a new location on that day. This allowed us to incorporate variance in the actual composition of travellers (infected versus non-infected), but because movement numbers were generated independently, it was possible for the number of infected people who stayed and the number who move in each patch to exceed or be fewer than the number of infected people in the patch. As we only wanted to incorporate variance into relative patterns of movement and not absolute numbers (particularly because the underlying values are proportions of people who moved and therefore cannot influence the total numbers of people infected), in any case in which the number of infected people who moved and the number who stayed differed from the total number of infected people in the origin patch, we rescaled values to the total number of infected people. Rescaling in this way meant the variance introduced by the Bernoulli trials could only influence relative movement patterns, and not actual numbers of infected people. Further, because we explicitly model the number of stayers in the same way as movers, rescaling should not introduce any bias in terms of the final relative movement patterns.

Through this model, stochasticity in the numbers and in the places with COVID-19 infections appears between simulation runs owing to variance in numbers of people becoming exposed, infectious and removed/recovered, as well as variance in numbers of people moving from one city to another. By modelling the COVID-19 epidemic in this way, we could simulate the incidence of COVID-19 cases, accounting for variance in recovery, infection and movement across many simulation runs (1,000). In addition, this allowed for us to account for uncertainty in contact rates after NPIs were implemented or lifted.

Simulation runs

Using this model, we quantified how the transmission of COVID-19 varied with different intervention scenarios and timings, as well as the potential of further transmission after the lifting of travel restrictions and contact distancing measures on 17 February 2020. As the earliest date of illness onset in cases was 2 December 2020 (ref. [3]), considering the underreporting of cases and the delay from infection to onset and identification of this novel virus, we started our simulations by infecting five people in Wuhan on 1 December 2019 and propagating the epidemic through time, varying factors including the timing and types of interventions used, assumed contact and recovery rates, and movement. We initially infected five people as a minimum number of infected people that prevented stochastic extinction of the epidemic during the initial days of simulation, and found no significant difference after three months, over simulation runs that started with three, five and eight people initially infected (though with three people initially infected, 50% of runs led to zero cases over the first week of simulation). When using data from other years we fixed the simulation dates around Chinese New Year and adjusted the start date of the epidemic accordingly.

The estimates of the model for the outbreak under NPIs as the baseline scenario were compared with reported COVID-19 cases across time and space. The sensitivity and specificity were also calculated to examine the performance of the model in predicting the occurrence of COVID-19 cases at the city level across China. The relative effects of NPIs were quantitatively assessed by comparing estimates of cases under various NPIs and timings with that of the baseline scenario. We also conducted a series of sensitivity analyses to understand the effect that changing epidemiological parameters had on the estimates and uncertainties of intervention efficacy. The software R v.3.6.1 (R Foundation for Statistical Computing) was used for data collation and analyses.

Ethical approval

Ethical clearance for collecting and using secondary data in this study was granted by the institutional review board of the University of Southampton (no. 48002). All data were supplied and analysed in an anonymous format, without access to personal identifying information.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data on the number of cases of COVID-19 reported by county, city and province across China are available from the data sources listed in the Supplementary Information, and the average days from illness onset to reporting of the first case by each county used in the modelling are detailed in Supplementary Table 1. The mobile phone datasets analysed during the current study are not publicly available as this would compromise the agreement with the data provider; however, information on the process of requesting access to the data that support the findings of this study is available from S.L., and the data on travel and contact reductions that were derived from the datasets and used in our model are detailed in Supplementary Table 2.

Code availability

The code for the model built in this study has been made openly available for further use at https://github.com/wpgp/BEARmod.

28. Wang, X., Liu, C., Mao, W., Hu, Z. & Gu, L. Tracing the largest seasonal migration on Earth. Preprint at https://arxiv.org/abs/1411.0983 (2014).
29. National Bureau of Statistics of China. *China Statistical Yearbook 2014* http://www.stats.gov.cn/english/Statisticaldata/AnnualData/ (2020).
30. Kraemer, M. U. G. et al. Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. *Nat. Microbiol.* **4**, 854–863 (2019).
31. Zhang, J. et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *Lancet Infect. Dis.* **20**, 793–802 (2020).

Acknowledgements We thank staff members at disease control institutions, hospitals and health administrations in areas across China in which outbreaks of COVID-19 occurred for field investigation, administration and data collection; Baidu Inc. for sharing population movement data; and Y. Zhu and S. Lai for collating online data. This study was supported by grants from the Bill and Melinda Gates Foundation (OPP1134076, OPP1195154); the European Union Horizon 2020 (MOOD 874850); the National Natural Science Fund of China (81773498); and the National Science and Technology Major Project of China (2016ZX10004222-009). A.J.T. is supported by funding from the Bill and Melinda Gates Foundation (OPP1106427, OPP1032350, OPP1134076, OPP1094793), the Clinton Health Access Initiative, the UK Department for International Development (DFID) and the Wellcome Trust (106866/Z/15/Z, 204613/Z/16/Z); H.Y. is supported by funding from the National Natural Science Fund for Distinguished Young Scholars of China (no. 81525023) and the Program of Shanghai Academic/Technology Research Leader (no. 18XD1400300); and O.P. is supported by funding from the National Science Foundation, USA (no. 1816075). The funders of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Author contributions S.L. designed the study, built the model, collected data, finalized the analysis, interpreted the findings and wrote the manuscript. N.W.R. built the model, analysed data, interpreted the findings and wrote the manuscript. L.Z. collected data, interpreted the

findings, commented on and revised drafts of the manuscript. J.R.F., O.P. and W.L. built the model, commented on and revised drafts of the manuscript. C.Z. collected data, interpreted the findings and commented on and revised drafts of the manuscript. A.J.T. interpreted the findings and revised drafts of the manuscript. A.W., M.S., X.D. and H.Y. interpreted the findings and commented on and revised drafts of the manuscript. All authors read and approved the final manuscript. The corresponding authors had full access to all of the data in the study and had final responsibility for the decision to submit for publication.

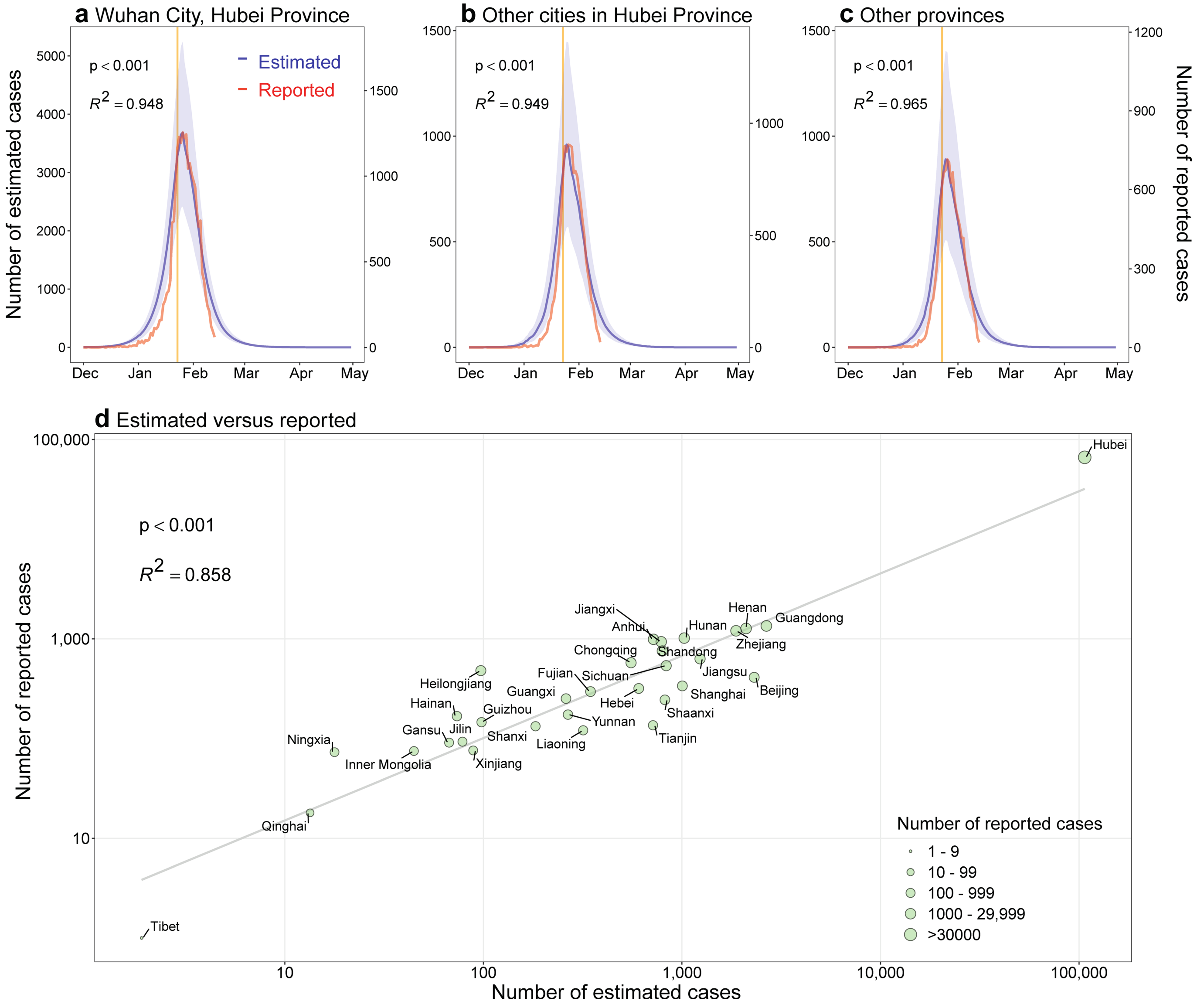
Competing interests H.Y. has received research funding from Sanofi Pasteur, GlaxoSmithKline, Yichang HEC Changjiang Pharmaceutical Company and Shanghai Roche Pharmaceutical Company. All other authors declare no competing interests.

Additional information

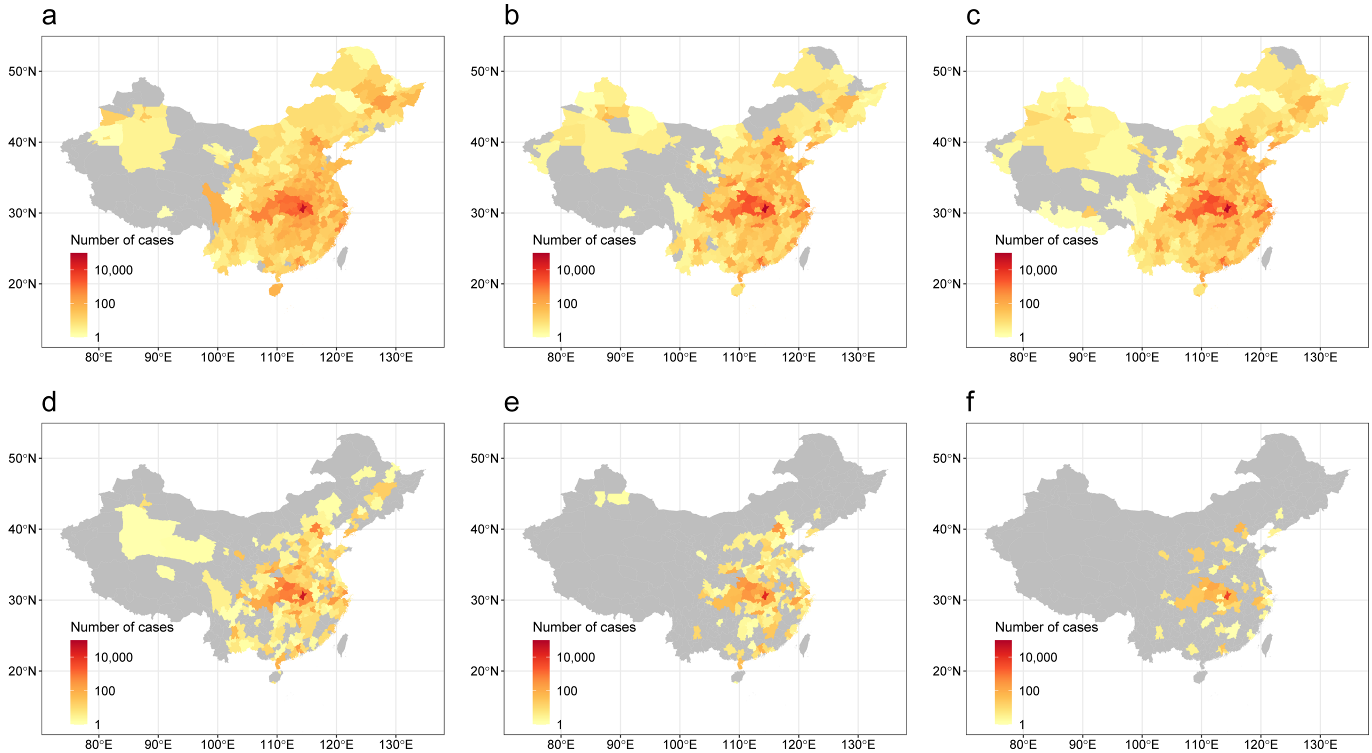
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2293-x.

Correspondence and requests for materials should be addressed to S.L., N.W.R. or A.J.T.

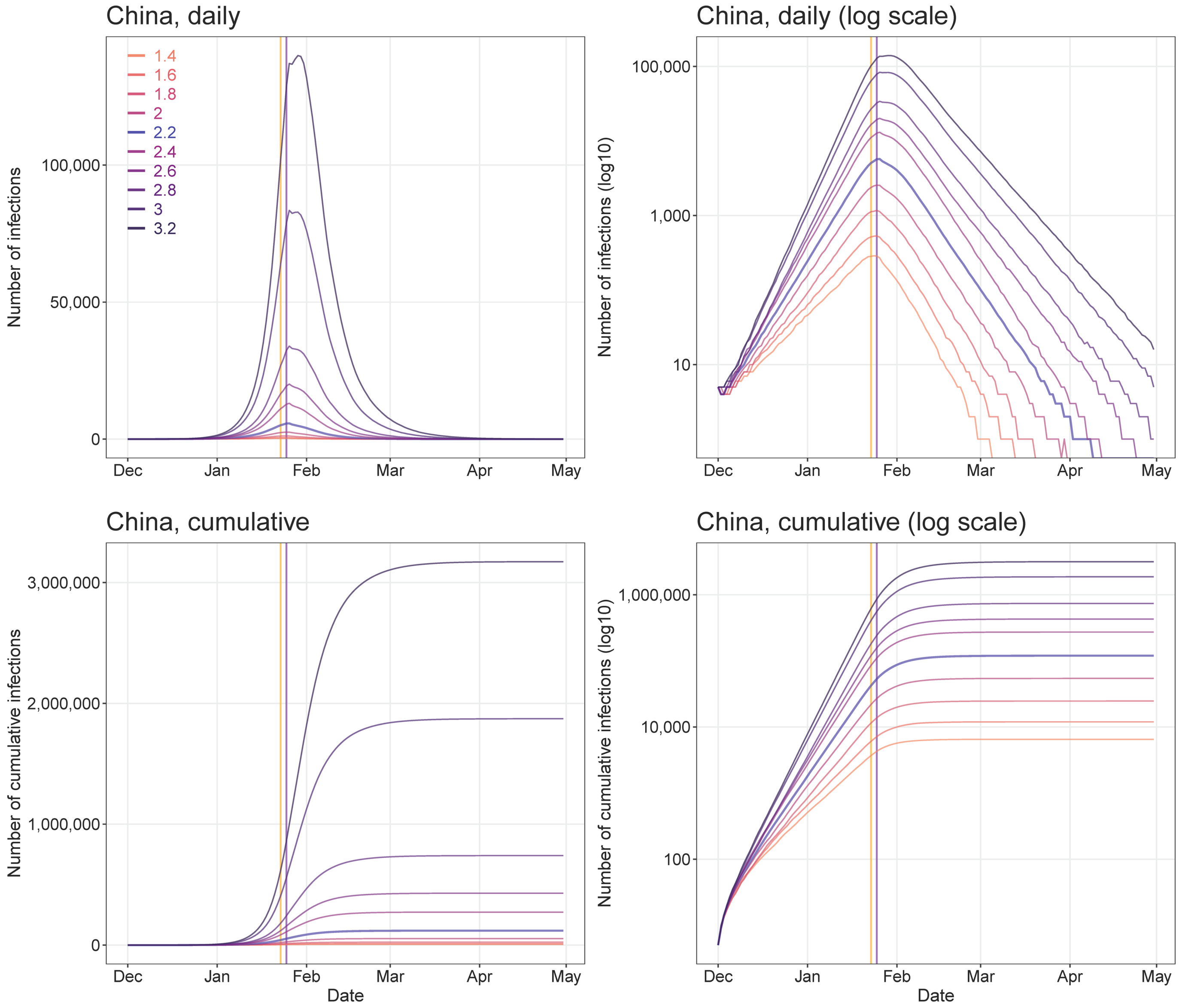
Peer review information *Nature* thanks Jukka-Pekka Onnela and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



Extended Data Fig. 1 | Estimated and reported epidemic curves of the COVID-19 outbreak in mainland China. **a**, The city of Wuhan in Hubei province. **b**, Other cities in Hubei province. **c**, Thirty other provincial regions in mainland China. The orange vertical lines indicate the date on which Wuhan's lockdown began (23 January 2020). The estimated epidemic curves of COVID-19 cases show the median (dark-blue line) and IQR (light-blue shading) of estimates (1,000 simulations), and the Pearson's correlation between the median of daily estimates and the number of daily reported cases by region as of 13 February 2020 is also shown. **d**, The Pearson's correlation between the total number of estimated cases and the total number of reported cases by province as of 29 February 2020. *P* value calculated by two-sided *t*-test.



Extended Data Fig. 2 | Areas affected by COVID-19 in mainland China under various intervention timings. a, A total of 308 cities reported COVID-19 cases, on the basis of data obtained from national and local health authorities, as of 29 February 2020. **b**, Affected areas (298 cities) estimated by models under interventions implemented at actual timing. **c**, Estimated affected areas (326 cities) under interventions implemented at actual timing, but without intercity travel restrictions. **d**, Estimated affected areas (192 cities) under interventions implemented one week earlier than actual timing. **e**, Estimated affected areas (130 cities) under interventions implemented two weeks earlier than actual timing. **f**, Estimated affected areas (61 cities) under interventions implemented three weeks earlier than actual timing. The administrative boundary maps were obtained from the National Platform of Common Geospatial Information Services of China (www.tianditu.gov.cn).



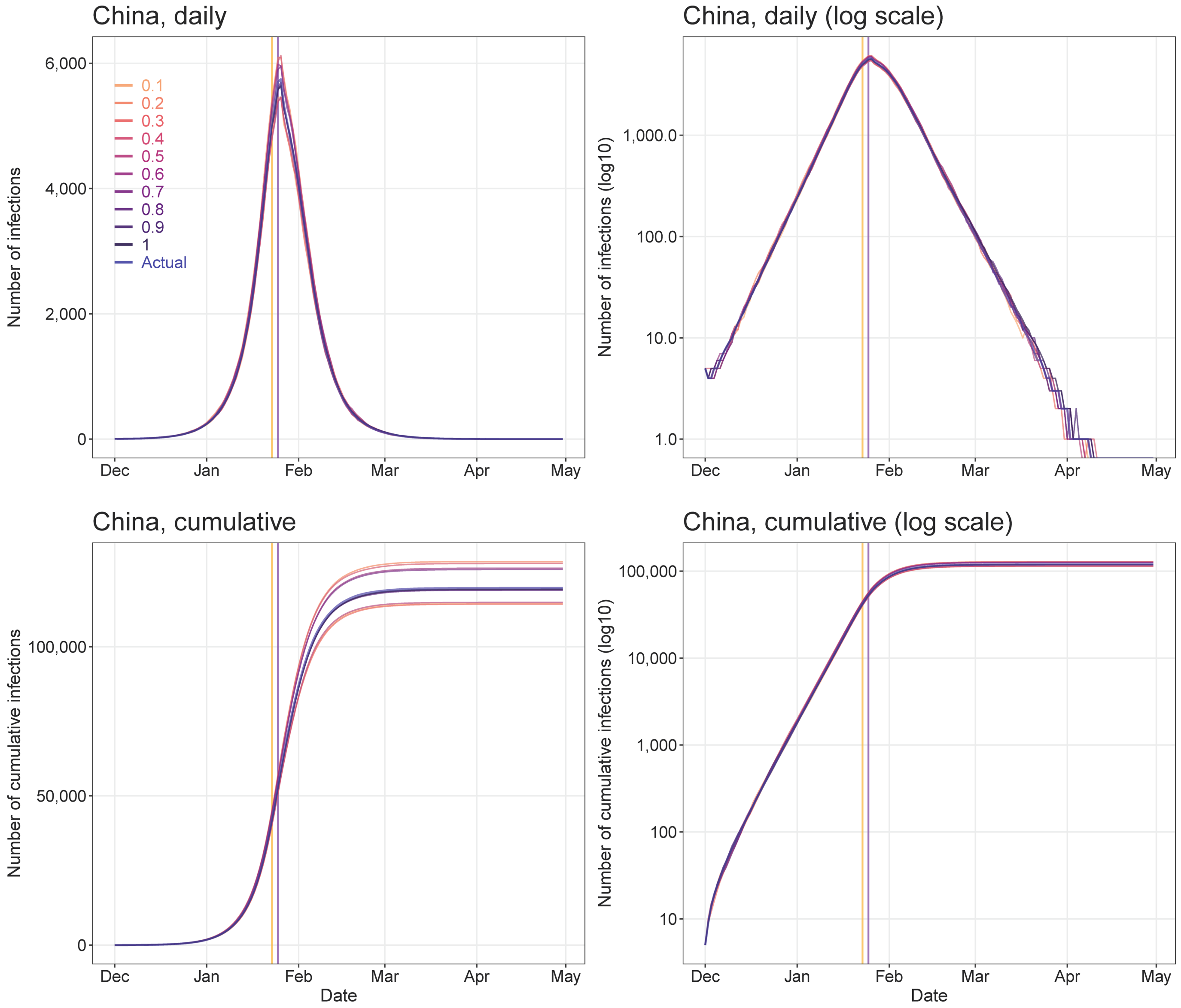
China, daily

China, daily (log scale)

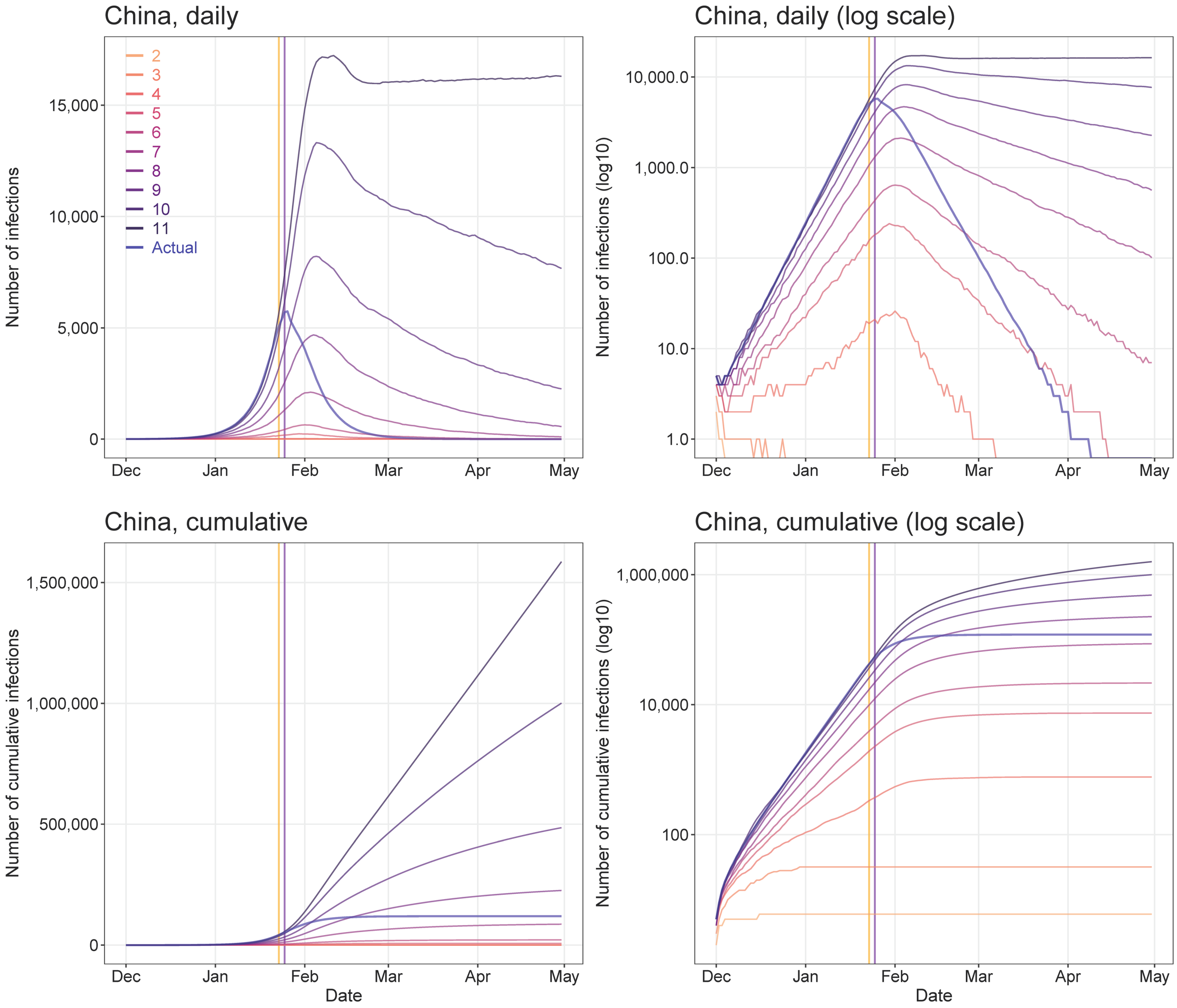
China, cumulative

China, cumulative (log scale)

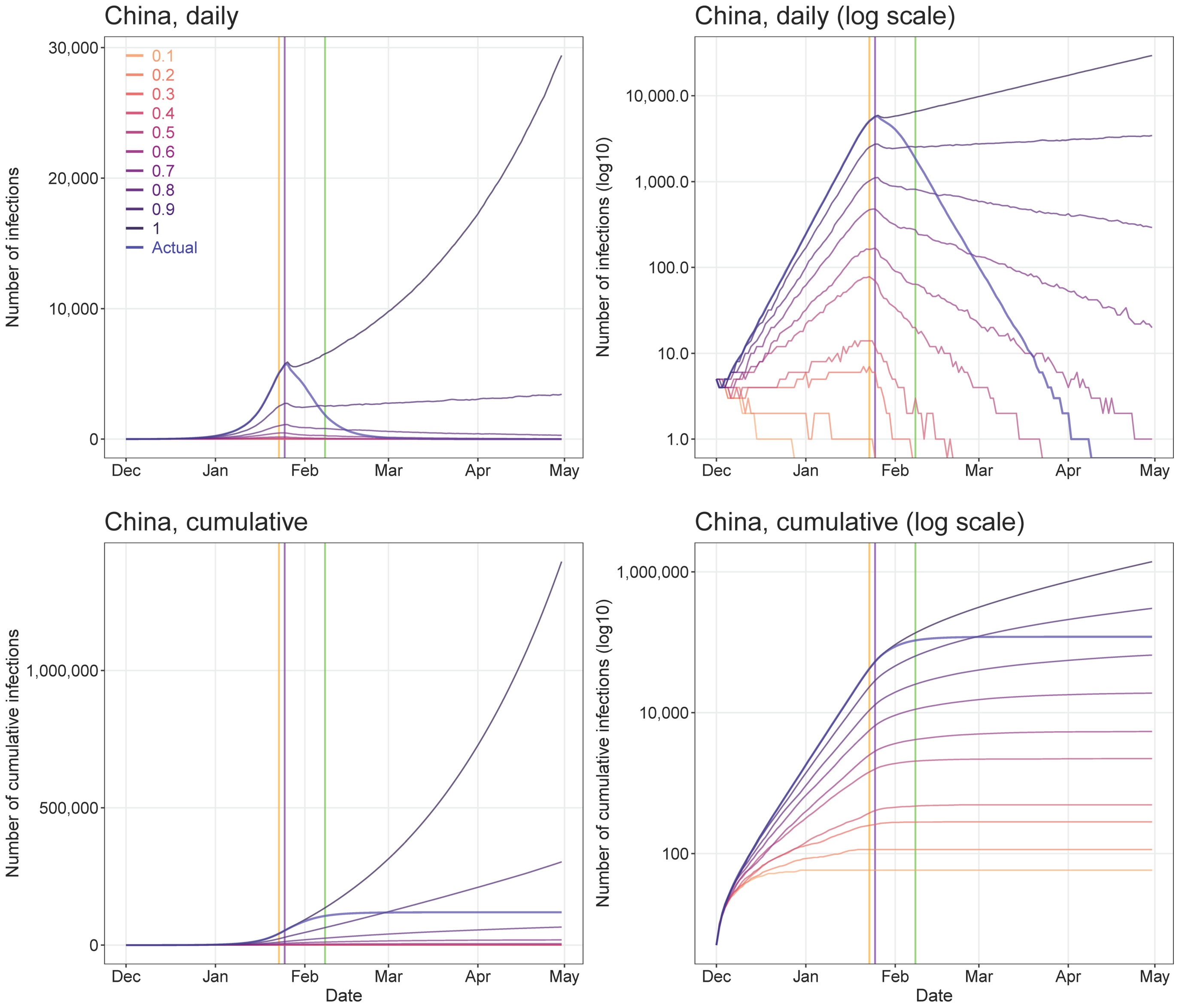
Extended Data Fig. 3 | Sensitivity of estimates of COVID-19 epidemics for various values of R_0. All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).



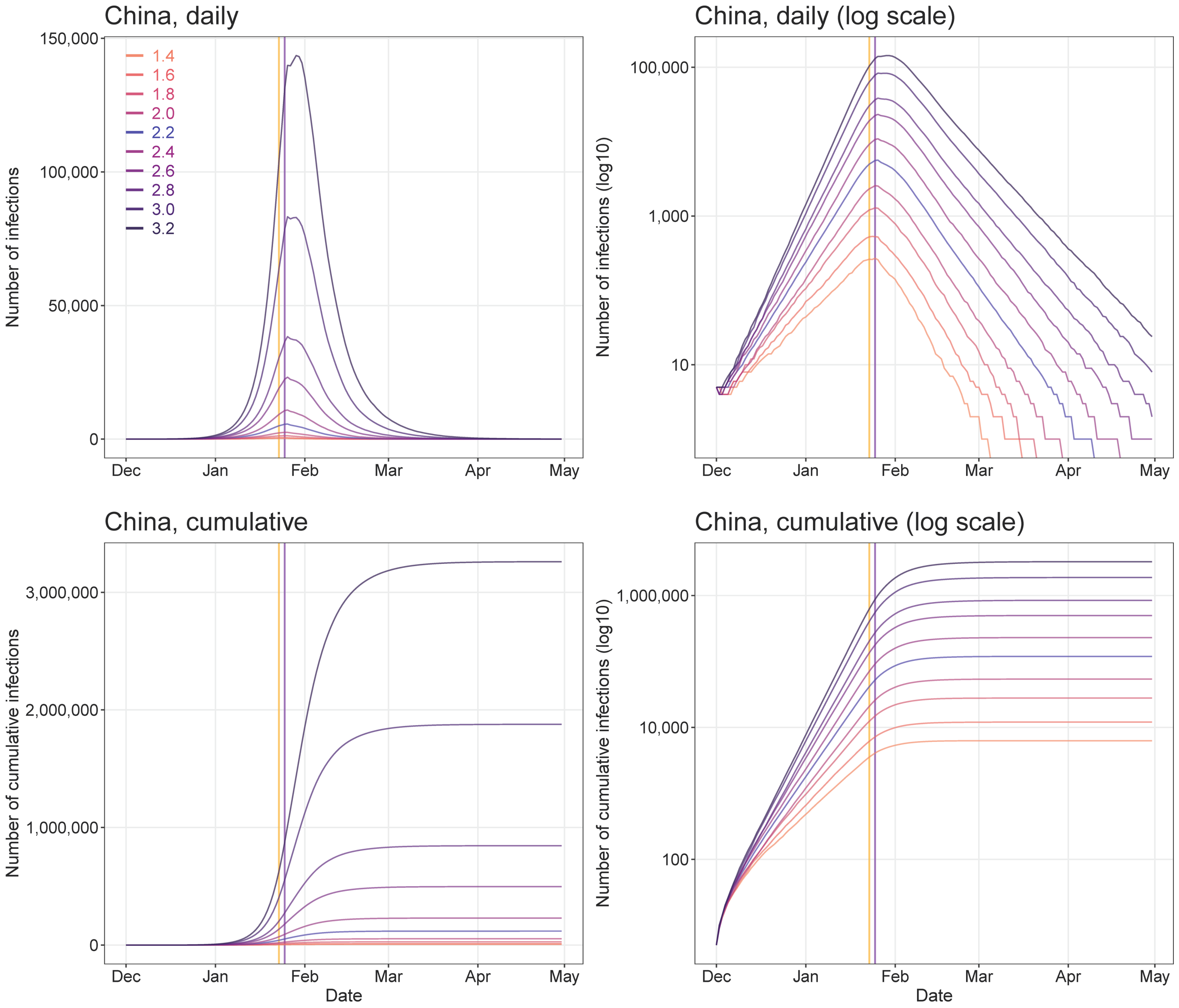
Extended Data Fig. 4 | Sensitivity of estimates of COVID-19 epidemics for various levels of intercity travel restrictions from 23 January 2020. All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. The actual percentages of intercity travel restrictions changed day by day across cities in China (0.1 indicates a 90% reduction from normal travel; 1 indicates no travel restrictions). Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).



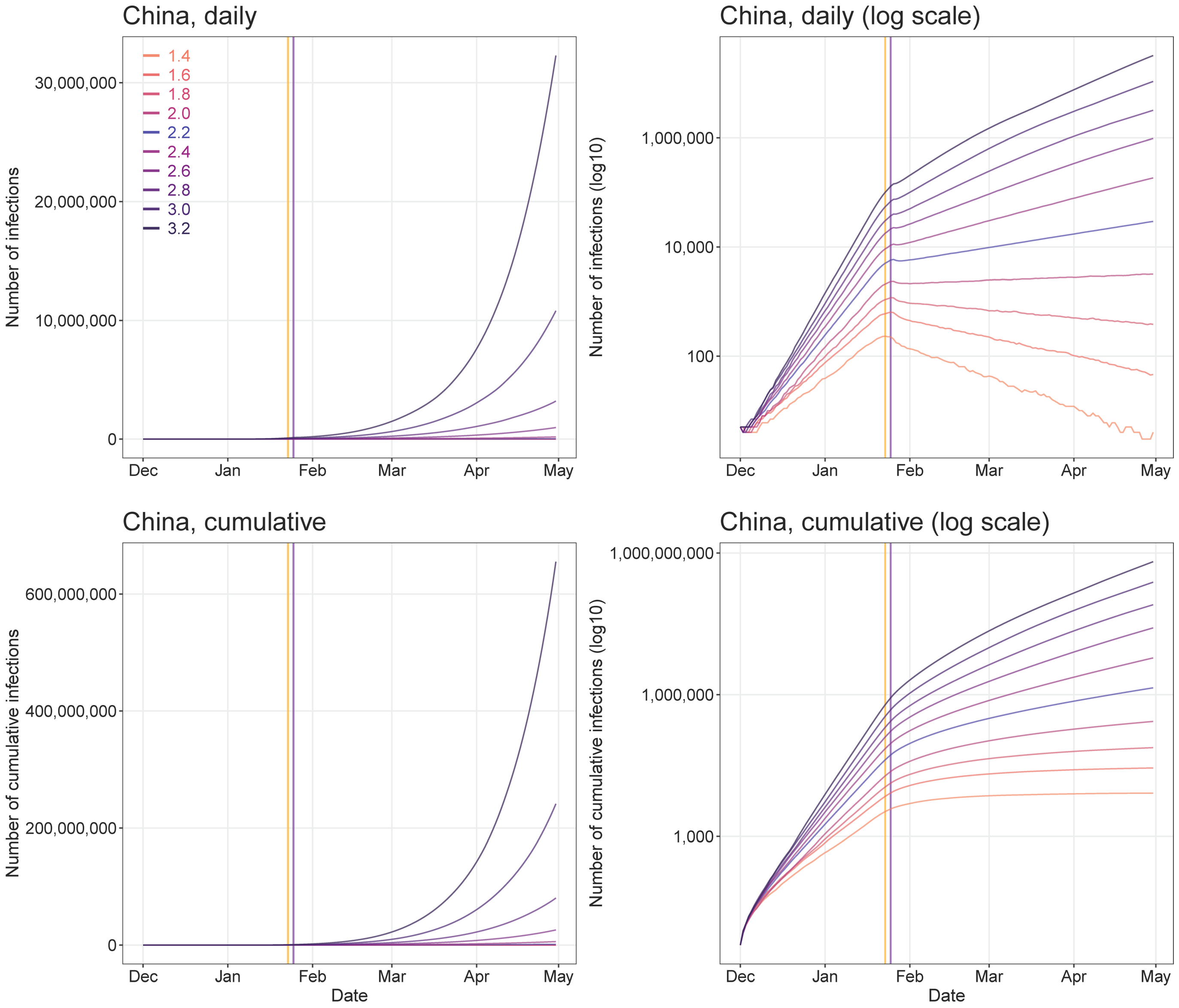
Extended Data Fig. 5 | Sensitivity of estimates of COVID-19 epidemics for various numbers of days from illness onset to report or isolation. All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. The actual delays of illness onset to report or isolation changed day by day (Supplementary Table 2). Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).



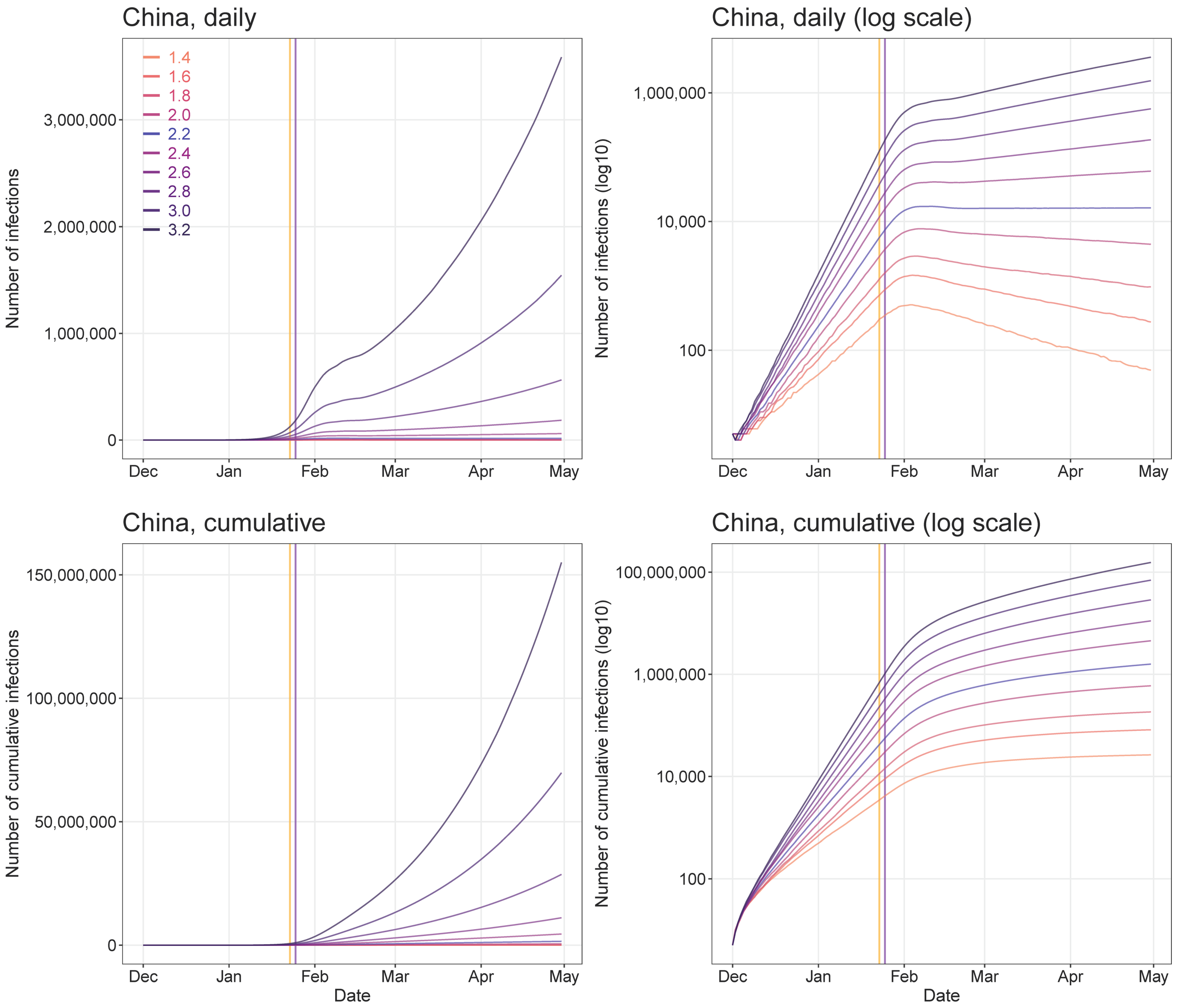
Extended Data Fig. 6 | Sensitivity of estimates of COVID-19 epidemics for various rates of contact. All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. The actual percentage of population contact (0.1 indicates 10% of usual contact, 1 means no contact restrictions) changed day by day across the country (Supplementary Table 1). Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).



Extended Data Fig. 7 | Sensitivity of estimates of COVID-19 epidemics for various values of R_0 and without intercity travel restrictions. All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).



Extended Data Fig. 8 | Sensitivity of estimates of COVID-19 epidemics for various values of R_0 and without contact restrictions within cities. All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).



Extended Data Fig. 9 | Sensitivity of estimates of COVID-19 epidemics for various values of R_0 and without improved timeliness of case detection and isolation. The delay from illness onset to detection and isolation was set as a constant of 11 days, which was the average delay from 16–18 January 2020. All other parameters, NPIs and input data were the same as the baseline model with $R_0 = 2.2$. Orange vertical lines indicate the date on which the lockdown of Wuhan began (23 January 2020); purple vertical lines indicate the date on which the Chinese New Year began (25 January 2020).

Extended Data Table 1 | Reports and estimates of COVID-19 cases in mainland China as of 29 February 2020

Interventions and timing	Wuhan City, Hubei Province	Other cities in Hubei Province	Other provinces	Mainland China
Under current non-pharmaceutical interventions (NPIs)				
No. of cases reported (%)[a]	49,122 (62)	17,785 (22)	12,917 (16)	79,824 (100)
Estimates of cases (%)	78,910 (69)	18,503 (16)	16,912 (15)	114,325 (100)
Interquartile range	51,952-111,280	11,029-28,685	9,499-27,033	76,776-164,576
Dates of estimated peak	Jan 25-27	Jan 24-26	Jan 24-26	Jan 25-27
Interval between NPIs and epidemic peak[b]	7 days	6 days	6 days	7 days
Percentage (%) of cases that could have been prevented with earlier interventions				
One week ahead	61 (45-79)	71 (55-86)	78 (62-90)	66 (50-82)
Two weeks ahead	84 (78-89)	90 (82-94)	91 (84-95)	86 (81-90)
Three weeks ahead	94 (92-96)	97 (95-99)	98 (97-99)	95 (93-97)
Estimated relative no. of cases with later interventions [c]				
One week delay	2.4 (1.6-3.5)	3.1 (1.8-4.6)	3.3 (2-5.4)	2.6 (1.8-3.8)
Two weeks delay	5.8 (4.0-8.6)	8.6 (5.3-12.8)	9.4 (6.1-14.6)	6.7 (4.6-10.0)
Three weeks delay	15.1 (9-21.1)	22.6 (13.5-33.9)	27.9 (17.5-42.8)	17.6 (11.2-25.5)
Estimated relative no. of cases under various NPIs [c]				
Without inter-city travel restriction	1.0 (0.6-1.3)	1.1 (0.7-1.7)	1.1 (0.7-1.7)	1.0 (0.6-1.4)
Without inner-city contact reduction	2.5 (1.7-3.7)	2.6 (1.5-4.2)	2.4 (1.2-4.0)	2.6 (1.7-3.7)
Without case early detection and isolation	5.0 (3.3-6.9)	5.6 (3.2-8.4)	5.1 (2.5-8.4)	5.0 (3.3-7.1)
Without all interventions above	51.4 (33.2-71.2)	91.6 (57.6-132.5)	124.7 (77.4-180)	67.3 (43.7-93.7)

[a] The reported data on COVID-19 cases were obtained from the Chinese National Health Commission as of 29 February 2020.

[b] The timeliness of case identification and reporting improved from 19 January 2020 and the travel restrictions and social distancing were implemented from 23 January 2020. We compared the peak dates by region with 19 January 2020 to define the interval from NPIs to epidemic peak.

[c] Referring to the median of estimates under actual interventions and timing.

The median and IQR of estimates are shown.

nature research

Corresponding author(s): Shengjie Lai

Last updated by author(s): Apr 17, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) was used to perform data collation and analyses.
Data analysis	R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) was used to perform data collation and analyses. The model built by this study has been made openly available for further use at https://github.com/wpgp/BEARmod.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data of COVID-19 cases reported by county, city, and province across China are available from data sources detailed in the Supplementary, and the average days from illness onset to report of the first case by each county used in the modelling are detailed in Supplementary Table 2. The mobile phone datasets analysed during the current study are not publicly available since this would compromise the agreement with the data provider, but the information on the process of requesting access to the data that support the findings of this study are available from Dr Shengjie Lai (Shengjie.Lai@soton.ac.uk), and the data of travel and contact reductions derived from the datasets and used in our model are detailed in Supplementary Table 1.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative observational and modelling study
Research sample	COVID-19 cases reported across mainland China as of February 29, 2020 were included in this study. As public awareness and enhanced case searching remained high throughout the study period, a high proportion of cases with symptoms was likely to have been detected, with nearly all reported cases eventually subjected to laboratory testing. However, the reported data of COVID-19 cases might not include asymptomatic and mild infections, and our model may have underestimated the total number of infections. The data on COVID-19 cases reported by county, city, and province across China are available from the data sources listed in the Supplementary Information File 3. This study also used population movement data across the country in 2020 and previous years, obtained from Baidu Location-based service. However, coverage biases of smartphone and Baidu users in population likely exist. Though a high percentage of the population owns smartphones in China, the mobile user group still does not cover specific subgroups of the population, particularly children. Therefore, our population movement data may provide an incomplete picture, and differences between the characteristics of smartphone owners and non-owners may also bias estimates in this study. Additionally, the magnitude and patterns of movements could change year by year.
Sampling strategy	This study included the numbers of all COVID-19 cases reported across mainland China as of February 29, 2020. Population movement data on human mobility of all Baidu users across the country were obtained from Baidu location-based service in 2014-2015 and 2020.
Data collection	We collated data of the first case reported by county across mainland China to measure the delay from illness to case report as a reference of the improved timeliness of case identification, isolation and reporting during the outbreak (Supplementary Information File 1). The daily number of COVID-19 cases by date of illness onset in Wuhan City, Hubei Province and other provinces as of February 13, 2020 were used to further validate the epicurves estimated in this study across time. The number of cases reported by city across mainland China as of February 29 were used to define the predictability of our model across space. These case data were collated from the websites of national and local health authorities, news media, and publications (Supplementary Information File 3). The epidemiological parameters estimated for the early stage of the outbreak in Wuhan from previous study (reference #5) were initially collected used to parameterise the epidemic before widely implementing interventions. Three population movement datasets, obtained from Baidu location-based services, were used in this study: 1) daily relative outbound and inbound flow of mobile phone users for each prefecture-level city (340 cities in mainland China) in 2020; 2) historical relative movement matrix with daily total number of users at city level from December 26, 2014 to May 26, 2015, aligning with the 2020 Chinese new year holiday period; 3) daily population movements at county level (2862 counties in China) from January 26 through April 30, 2014, aligning with the 2015 and 2020 Chinese new year holiday period.
Timing	COVID-19 cases: December 2, 2019 - February 29, 2020. Three Baidu population movement datasets: 1) January 26, 2014- April 30, 2014; 2) December 26, 2014 - May 26, 2015; 3) January 1, 2020 - April 13, 2020.
Data exclusions	Before conducting this study, we already noticed that there was an abnormal increase of cases in Wuhan City and Hubei Province on February 1, 2020, based on the date of illness onset. The case definition has been adjusted and a large number of clinically diagnosed cases before laboratory confirmation have been retrospectively reported into the information system since 12 February. However, the spike on February 1 might not represent the actual infection patterns. We have discussed this issue and underlying causes (e.g. changes of definitions, reporting delay, system error, incorrect reporting of the onset date) with epidemiologists in China, but exact reasons remain unclear. Therefore, before comparing reported data with estimates in our study, we interpolated the number on February 1 by using the mean of the numbers of cases reported on January 31 and February 2 in the epicurves of Wuhan and Hubei Province.
Non-participation	As this study collected and used secondary data from disease surveillance and Baidu location-based service, we did not access to the raw data and we don't know how many participants dropped out/declined participation. However, the number of COVID-19 cases might be unreported as asymptomatic and mild infections exist. The mobile user group does not cover specific subgroups of the population, particularly children, and not all mobile owners use the Baidu location-based service. Therefore, our population movement data may provide an incomplete picture of movement of all population in China, and the spatiotemporal and demographic variations in the behaviour of phone users could have biased population distribution and travel estimates.
Randomization	We did not randomly sample COVID-19 cases and movement data from population. In our SEIR modelling framework, we conducted 1000 simulations to account for the uncertainty of estimates.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Envelope protein ubiquitination drives entry and pathogenesis of Zika virus

https://doi.org/10.1038/s41586-020-2457-8

Received: 22 September 2018

Accepted: 27 April 2020

Maria I. Giraldo[1,2,18], Hongjie Xia[3,18], Leopoldo Aguilera-Aguirre[1], Adam Hage[1], Sarah van Tol[1], Chao Shan[3], Xuping Xie[3], Gail L. Sturdevant[4], Shelly J. Robertson[4], Kristin L. McNally[4], Kimberly Meade-White[4], Sasha R. Azar[1,5,6], Shannan L. Rossi[6,7], Wendy Maury[8], Michael Woodson[9], Holly Ramage[10], Jeffrey R. Johnson[11,12,13,17], Nevan J. Krogan[11,12,13,14], Marc C. Morais[3,9], Sonja M. Best[4], Pei-Yong Shi[3,6,9,15,16]✉ & Ricardo Rajsbaum[1,6]✉

Zika virus (ZIKV) belongs to the family *Flaviviridae*, and is related to other viruses that cause human diseases. Unlike other flaviviruses, ZIKV infection can cause congenital neurological disorders and replicates efficiently in reproductive tissues[1–3]. Here we show that the envelope protein (E) of ZIKV is polyubiquitinated by the E3 ubiquitin ligase TRIM7 through Lys63 (K63)-linked polyubiquitination. Accordingly, ZIKV replicates less efficiently in the brain and reproductive tissues of *Trim7*$^{-/-}$ mice. Ubiquitinated E is present on infectious virions of ZIKV when they are released from specific cell types, and enhances virus attachment and entry into cells. Specifically, K63-linked polyubiquitin chains directly interact with the TIM1 (also known as HAVCR1) receptor of host cells, which enhances virus entry in cells as well as in brain tissue in vivo. Recombinant ZIKV mutants that lack ubiquitination are attenuated in human cells and in wild-type mice, but not in live mosquitoes. Monoclonal antibodies against K63-linked polyubiquitin specifically neutralize ZIKV and reduce viraemia in mice. Our results demonstrate that the ubiquitination of ZIKV E is an important determinant of virus entry, tropism and pathogenesis.

ZIKV is transmitted primarily by peridomestic *Aedes* mosquitoes, but also can be acquired through sexual, vertical and blood transfusion routes[1,2]. ZIKV infection causes congenital abnormalities in fetuses of pregnant women infected with the virus[3]. Although ZIKV is closely related to other flaviviruses that cause human diseases (including dengue virus (DENV), West Nile virus (WNV) and yellow fever virus), the mechanism of how ZIKV causes neurological disorders or replicates in reproductive tissues remains unclear.

The ubiquitination of proteins is a post-translational modification process with many cellular functions, including the regulation of virus replication[4]. There is previous evidence that flaviviruses use the host ubiquitination system for replication[5–7]; however, whether flaviviruses carry ubiquitin in the infectious virion or whether the ubiquitination machinery is involved in determining virus tropism and pathogenesis has not been explored. Tripartite motif (TRIM) proteins are a large family of E3 ubiquitin ligases that mediate the transfer of ubiquitin to target proteins, and many of these ligases are known to inhibit viral replication[4,8,9]. However, very few examples exist of TRIM proteins being exploited by viruses to promote virus replication[9,10]. Here we report that the E of ZIKV is ubiquitinated by the E3 ubiquitin ligase TRIM7, and that this modification is a determinant of tissue tropism. A proportion of virions contain ubiquitinated E, which promotes more-efficient attachment and entry into host cells.

Flavivirus E is ubiquitinated

Previous studies have shown that proteasome inhibitors reduce the replication of DENV[7,11–13]. Consistent with this, JEG-3 cells (a cell line derived from human placenta) that are pretreated with the proteasome inhibitor MG132 are more resistant to ZIKV infection (Extended Data Fig. 1a). To examine whether the ubiquitination of viral proteins has a role in flavivirus biology, we performed mass spectrometry analysis of samples from cells infected with West Nile virus, DENV-2 or ZIKV. This analysis identified ubiquitination on the K38 residue, which is conserved among flaviviruses (Extended Data Fig. 1b). Another ubiquitination site on K281 at the hinge region (known as the 'kl loop') of the E of ZIKV was identified; however, K281 is not conserved in flaviviruses[14] (Extended Data Fig. 1b). We focused our studies on E because of the essential function of this protein in virus entry[15]. Co-immunoprecipitation assays with Huh7 cells infected with DENV or ZIKV confirmed that E was ubiquitinated in

[1]Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, TX, USA. [2]Centro de Investigaciones Biomédicas, Universidad del Quindío, Armenia, Colombia. [3]Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, USA. [4]Laboratory of Virology, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT, USA. [5]Institute for Translational Sciences, University of Texas Medical Branch, Galveston, TX, USA. [6]Institute for Human Infections and Immunity, University of Texas Medical Branch, Galveston, TX, USA. [7]Department of Pathology, University of Texas Medical Branch, Galveston, TX, USA. [8]Department of Microbiology and Immunology, University of Iowa, Iowa City, IA, USA. [9]Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, TX, USA. [10]Department of Microbiology, University of Pennsylvania, Philadelphia, PA, USA. [11]Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA, USA. [12]Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA, USA. [13]Gladstone Institute for Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA, USA. [14]Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [15]Sealy Institute for Vaccine Sciences, University of Texas Medical Branch, Galveston, TX, USA. [16]Department of Pharmacology and Toxicology, University of Texas Medical Branch, Galveston, TX, USA. [17]Present address: Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [18]These authors contributed equally: Maria I. Giraldo, Hongjie Xia. ✉e-mail: peshi@utmb.edu; rirajsba@utmb.edu

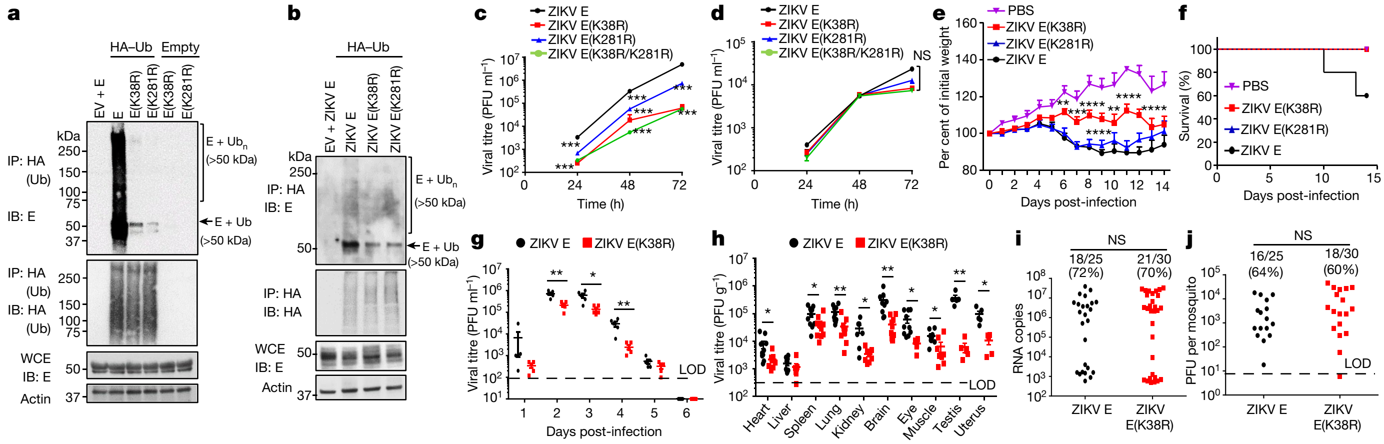


Fig. 1 | ZIKV-E ubiquitination on K38 and K281 promotes virus replication in cells and in vivo. a, Whole-cell extracts (WCE) from HEK293T cells transfected with empty vector (EV), wild-type ZIKV E or ZIKV mutants and HA–Ub were used for immunoprecipitation (IP) with anti-HA beads. **b**, JEG-3 cells that stably express HA–Ub were infected with wild-type ZIKV E, or ZIKV mutants followed by HA immunoprecipitation. Because the mutant viruses are attenuated, the input E was normalized for immunoprecipitation. **c**, **d**, Virus titres in supernatants from infected JEG-3 cells (**c**) or mosquito C6/36 cells (**d**), at a multiplicity of infection (MOI) of 0.5. Representatives from 2 independent experiments; $n = 3$ technical replicates, mean ± s.e.m., $^{***}P < 0.001$. **e**–**h**, A129 mice mock-infected ($n = 5$) or infected with ZIKV mutants (1×10^4 plaque-forming units (PFU); 9 mice per group, from 2 independent experiments). **e**, Body weight. Two-way analysis of variance (ANOVA), Tukey's test. PBS, phosphate-buffered saline (used as control). **f**, Survival. Virus titres are shown in Extended Data Fig. 3c, d. **g**, **h**, Same experiment as in **e** repeated with 5 male and 5 female mice (viraemia (serum titres)) (**g**) and organ viral titres (day 6 post-infection) (**h**). Unpaired two-sided *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$. **i**, **j**, Mosquito infectivity. *Aedes aegypti* mosquitoes were fed with a blood meal (10^6 PFU ml^{-1}) of ZIKV. At day 10, individual mosquitoes were quantified for viral RNA (by qPCR) (**i**) and virus (by plaque assay) (**j**). LOD, limit of detection. Unpaired, two-sided *t*-test; NS, not significant ($P > 0.05$).

both viruses (Extended Data Fig. 1c). Examination of the type of ubiquitin linkage revealed that, in ZIKV, ubiquitinated E was mostly associated with K63-linked polyubiquitin chains (Extended Data Fig. 1d). We also found that proteasome inhibition significantly reduced viral RNA replication at later time points, but had no effects on virus entry and/or uncoating (Extended Data Fig. 1e)—as has previously been proposed for DENV[5,6]. Because E is critical in mediating virus entry and because proteasome inhibition does not have an effect early during infection, we focused our studies on the role of K63-linked polyubiquitination of E independent of the proteasome at early steps of the viral infection cycle.

Importance of ubiquitination on E K38 and K281

To test whether ZIKV is ubiquitinated on the K38 residue and to further confirm ubiquitination on K281, we performed co-immunoprecipitation assays of ubiquitin fused to a haemagglutinin peptide tag (HA–Ub) in the presence of wild-type E or K-to-R mutants on residues K38 and K281 (E(K38R) and E(K281R), respectively). We found that the ubiquitination of E was substantially reduced on E(K38R) and E(K281R) mutants, which confirms that E is ubiquitinated on both residues (Fig. 1a). On the basis of the molecular weights of ubiquitin (about 8.5 kDa) and E (about 48 kDa), a proportion of ubiquitinated E appears to be in the form of mono- or di-ubiquitinated E, or conjugated to a mix of larger polyubiquitin chains (a smear of over 50 kDa) (Fig. 1a). To examine the functional relevance of ubiquitination in the context of infectious ZIKV, we generated recombinant viruses that lack ubiquitination on E (E(K38R)-mutant ZIKV, E(K281R)-mutant ZIKV, or E(K38R/K281R)-mutant ZIKV). Co-immunoprecipitation assays confirmed reduced ubiquitination on the E(K38R)- and E(K281R)-mutant viruses (Extended Data Fig. 2a; co-immunoprecipitation was normalized to equal levels of input E in infected cells, as shown in Fig. 1b). Compared with wild-type ZIKV, both the E(K38R)- and E(K281R)-mutant viruses were highly attenuated in JEG-3 cells as well as in another cell line (HTR-8) derived from human placenta (Fig. 1c, Extended Data Fig. 2b, c). However, the replication level of E(K38R)-mutant, but not E(K281R)-mutant, ZIKV was also significantly reduced in testis (15P-1) and liver (Huh7) cells (Extended Data Fig. 2d, e). The input dose of wild-type and mutant viruses was equal (Extended Data Fig. 2f). By contrast, wild-type and mutant viruses replicated to similar levels in mosquito C6/36 cells (Fig. 1d). The E(K38R/K281R)-mutant ZIKV did not show an additive effect and replicated in a manner similar to that of the E(K38R)-mutant virus (Fig. 1c, d). Therefore, the ubiquitination of E has an important role in viral replication in the human, but not the mosquito, host.

Effects of lack of E ubiquitination in vivo

Because ubiquitination of wild-type E is detected in mouse brains and testis (Extended Data Fig. 3a, b) (two major sites of ZIKV replication during in vivo infection[16]), we examined whether a lack of ubiquitination on E would lead to altered tissue tropism and—consequently—pathogenesis using a previously established mouse model for ZIKV infection (*Ifnar1*$^{-/-}$, A129 mice)[16]. Consistent with the cell culture results, the E(K38R)-mutant ZIKV was significantly attenuated in vivo (Fig. 1e, f, Extended Data Fig. 3c, d). Infection with wild-type ZIKV resulted in weight loss, and death rates of 40%. By contrast, mice infected with the E(K38R)-mutant ZIKV showed significantly less weight loss (Fig. 1e) with no major signs of disease or death (Fig. 1f). Accordingly, viral titres of E(K38R)-mutant ZIKV in serum (on day 2) and in the brain and testis (on day 8) were significantly lower than wild-type ZIKV (Extended Data Fig. 3c, d). Although infection with the E(K281R)-mutant ZIKV did not show significant differences overall as compared to wild-type virus, it caused slightly less weight loss—which resulted in 100% survival. The difference in virus titres between wild-type and mutant viruses in the eye was marginal compared to brain and testis, even though the eye is another target of ZIKV infection[17]. To explore the possibility of differential tropism between wild-type ZIKV and E(K38R)-mutant ZIKV, we repeated the experiment and measured viral titres in additional tissues. Consistent with the data described above, the level of viraemia with the E(K38R)-mutant ZIKV was significantly lower than that with wild-type ZIKV (Fig. 1g). The largest reduction in viral titres between infection with E(K38R)-mutant and wild-type ZIKV (about 1 to 2 log) was observed in brain and reproductive tissues (uterus and testis); smaller differences (about 2–5-fold) were found in heart, spleen, lung, kidney, eye and muscle tissues (Fig. 1h). By contrast, comparable infection rates

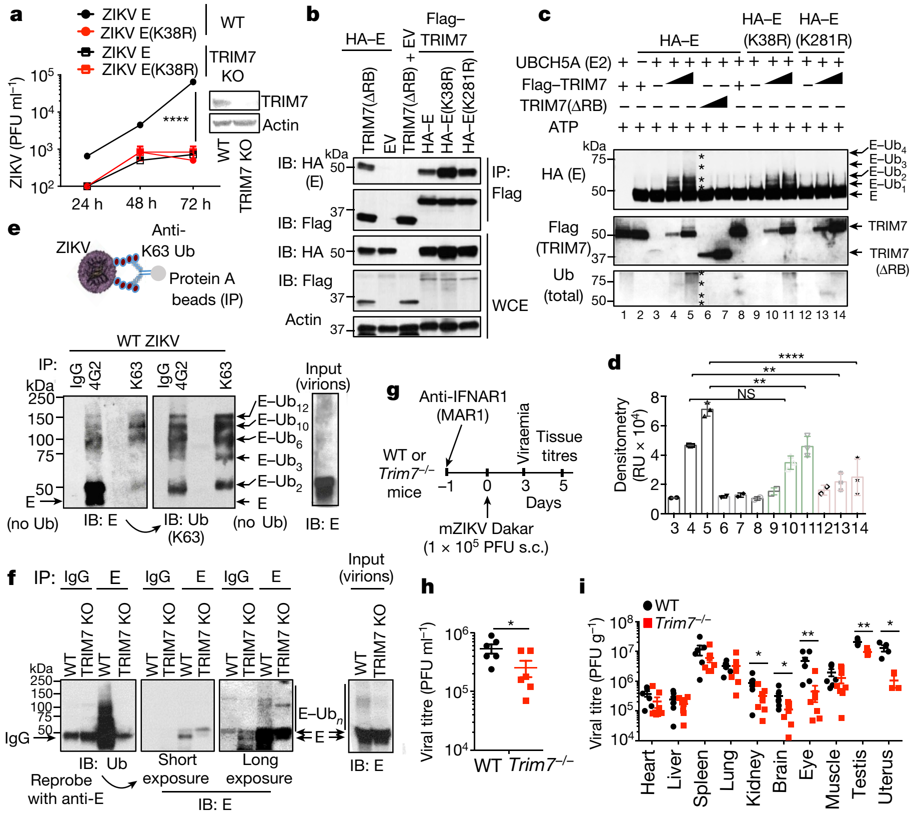


Fig. 2 | TRIM7 ubiquitinates ZIKV E and promotes virus replication. **a**, Virus titres from TRIM7-knockout (TRIM7 KO) JEG-3 cells infected with wild-type and mutant ZIKV (MOI 0.5). $n = 3$ technical replicates, mean ± s.e.m., multiple *t*-test, Holm–Sidak correction, $^{****}P < 0.0001$. WT, wild type. **b**, HEK293T cells transfected with TRIM7 or a short isoform (TRIM7(ΔRB)), wild-type or mutant E, followed by immunoprecipitation. Representative of two independent experiments. **c**, TRIM7 ubiquitinates recombinant E on both K38 and K281 in an in vitro ubiquitination assay. Representative of four independent experiments. **d**, Densitometry for experiment shown in **c**. $n = 3$, mean ± s.e.m., one-way ANOVA, Tukey's multiple comparison, $^{**}P < 0.01$, $^{****}P < 0.0001$, NS, not significant ($P > 0.05$). RU, relative units. **e**, K63-linked polyubiquitinated E was detected on ZIKV particles concentrated from supernatants from Vero cells after immunoprecipitation with anti-K63-linked ubiquitin or anti-E 4G2 antibodies. A diagram for the K63-linked immunoprecipitation of viruses is shown at the top of **e**. After immunoblot (IB) with anti-E, the blots were reprobed with anti-K63-linked ubiquitin. **f**, Immunoprecipitation with anti-E (4G2) of virus stocks from wild-type or TRIM7-knockout JEG-3 cells. Representative of two independent experiments. **g**–**i**, *Trim7*$^{-/-}$ and *Trim7*$^{+/+}$ littermate mice from three separate CRISPR knockout lines (4–5 weeks old) were treated intraperitoneally with anti-IFNAR1 (MAR1-5A3). The next day, infections were performed with the mouse-adapted (m)ZIKV Dakar strain (10^5 PFU injected subcutaneously (s.c.) into the foot pad). $n = 7$ *Trim7*$^{+/+}$ and 8 *Trim7*$^{-/-}$ mice. **h**, **i**, Serum (**h**) and tissue titres (**i**). Unpaired two-sided *t*-test. $^{*}P < 0.05$, $^{**}P < 0.01$.

between the wild-type and E(K38R)-mutant ZIKV were detected using quantitative reverse-transcription PCR (Fig. 1i) and plaque assay (Fig. 1j) in *Aedes aegypti* mosquitos at day 10 after a blood meal. Together, these data suggest that a lack of ubiquitination specifically on the K38 residue of E reduces viral pathogenesis and virus replication in a tissue-specific manner in the mammalian—but not in the mosquito—host, and that the ubiquitination of E may have a role in tissue and species tropism.

TRIM7 ubiquitinates E to promote replication

Because our data indicate that ubiquitination of the ZIKV E promotes viral replication and that ubiquitination of E may be a conserved feature in flaviviruses, we searched the literature for E3 ubiquitin ligases that had previously been reported to promote flavivirus replication in genome-wide short interfering (si)RNA knockdown studies. TRIM7 (also known as GNIP[18]), a member of the TRIM family of E3 ligases[4,9], has previously been identified as a potential proviral factor for yellow fever virus (supplementary table 1 of ref. [19]). Expression of the full-length TRIM7 isoform can be detected in the known sites of ZIKV replication[16], including the placenta[20], brain and testis—although it is also expressed at different levels in other tissues[21] (Extended Data Fig. 4a). When TRIM7 expression was knocked down via siRNA, ZIKV replication was significantly reduced in JEG-3 cells as well as in brain-derived HTB-15 cells (Extended Data Fig. 4b, c). These proviral effects of TRIM7 require an intact K38 residue on E, because—although wild-type ZIKV replicates to lower levels in JEG-3 cells in which TRIM7 is knocked out using CRISPR—no additional difference was observed upon infection with E(K38R)-mutant ZIKV between wild-type and TRIM7-knockout cells (Fig. 2a). In contrast to ZIKV, deletion of TRIM7 did not affect DENV replication in a lung A549 TRIM7-knockout cell line (Extended Data Fig. 4d). TRIM7 has previously been proposed to have antiviral roles against norovirus[22], potentially via induction of IFN[23]; indeed, TRIM7-knockout cells have reduced IFNβ induction upon ZIKV infection or stimulation with the double-stranded RNA mimic polyinosinic–polycytidylic acid (poly(I:C)) (Extended Data Fig. 4e, f). However, our data suggest that the proviral roles of TRIM7 are dominant over its potential IFN-mediated antiviral roles. Furthermore, ectopically expressed TRIM7 increased K63-linked polyubiquitination of wild-type E but not E(K38R) or E(K281R) (Extended Data Fig. 4g) and correlated with increased virus titres when overexpressed in the Huh7 cell line (Extended Data Fig. 4h), in which ZIKV does not normally replicate to optimal levels. Consistent with these data, full-length TRIM7, as well as its short isoform (which lacks the RING–BBOX domains but retains the B30.2 domain), interact with wild-type E, E(K38R) and E(K281R) (Fig. 2b). Endogenous TRIM7 also co-immunoprecipitated with E in cells infected with ZIKV (Extended Data Fig. 4i), which confirms the interaction between TRIM7 and E. Finally, TRIM7 together with the E2 conjugating enzyme UBCH5A, which has previously been identified as interacting with and promoting TRIM7-mediated K63-linked ubiquitination[24,25], directly ubiquitinated recombinant ZIKV E on both K38 and K281 in an in vitro ubiquitination assay (Fig. 2c; quantification shown in Fig. 2d). In this in vitro system, although ubiquitination by TRIM7 is reduced in E(K38R), it appears that some compensation on other residues can occur under these conditions (compare lanes 10 and 11 with 4 and 5 in Fig. 2c, d).

Because TRIM7 has previously been suggested to localize in the Golgi[18], we hypothesized that TRIM7 may be recruited to intracellular membranes during exocytosis of progeny virus, where it could ubiquitinate E. In non-infected cells, TRIM7 showed weak diffuse staining with some apparent localization in cytoplasmic structures, as has previously been reported[8,18]. In addition, a low proportion of TRIM7 colocalized with wheat germ agglutinin, a lectin dye that labels glycoconjugates enriched in Golgi[26] (Extended Data Fig. 5a, b middle). However, upon ZIKV infection there was a notable reorganization of intracellular membranes, as has previously been reported during flavivirus infection[27]. Furthermore, ZIKV infection relocalized TRIM7 to these membranes, where a small proportion colocalized with E (Extended Data Fig. 5a, b top). Cell fractionation also showed both TRIM7 and its E2 conjugating enzyme UBCH5A cofractioned with the reticulum marker calnexin in infected cells (Extended Data Fig. 5c).

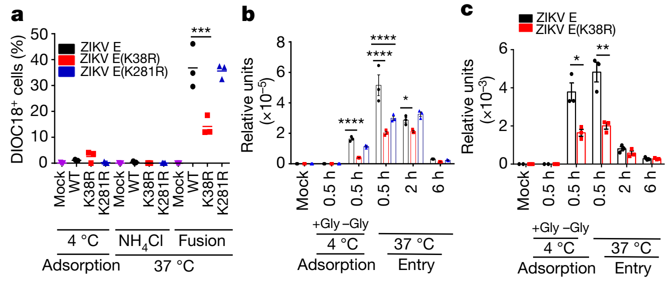


Fig. 3 | Ubiquitination of ZIKV E promotes virus attachment and fusion of the virus and the endosome membrane. a, Virus–endosome fusion. Different forms of ZIKV were labelled with DiOC18. After filtration, viruses were incubated at 4 °C with JEG-3 cells at MOI 2. After 30 min, cells were washed and collected for quantification. Additional samples were then incubated at 37 °C for 1 h, with or without NH_4Cl to block acidification (as control), washed, fixed and quantified by fluorescence-activated cell sorting. **b**, **c**, Viral RNA (qPCR). Ubiquitination of E promotes virus attachment. JEG-3 cells (**b**) or human primary induced pluripotent neural stem cells (**c**) were incubated with viruses for the indicated times as described in **a**. Incubation at 4 °C without glycine (−Gly) represents attached viruses. Each of the panels is representative of two independent experiments. $n = 3$ technical replicates, mean ± s.e.m. Unpaired two-sided *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

Ubiquitinated E in infectious ZIKV particles

We next tested whether mature ZIKV particles released from infected cells contained ubiquitinated E. Supernatants collected from ZIKV-infected JEG-3 cells showed detectable levels of ubiquitinated E (Extended Data Fig. 6a). Moreover, K63-linked polyubiquitinated E was detected on ZIKV particles concentrated from supernatants from Vero cells after immunoprecipitation with an anti-K63-linked-ubiquitin-specific antibody, and was also able to detect ubiquitinated E when virions were isolated using an anti-E (4G2) antibody (Fig. 2e). On the basis of molecular weight, potentially up to 12 ubiquitin molecules covalently attached to E could be detected (Fig. 2e). In addition, ubiquitinated E was also detected from virus stocks grown and concentrated from wild-type JEG-3 cells, but was strongly reduced in virus grown in TRIM7-knockout JEG-3 cells (Fig. 2f). Although deletion of TRIM7 in A549 cells also reduced the infectivity of progeny virus (Extended Data Fig. 6b), it did not affect viral RNA replication or virion release (Extended Data Fig. 6c, d). K63-linked polyubiquitinated E was also detected after immunoprecipitation of E from supernatants containing wild-type ZIKV, and was reduced in E(K38R)- and E(K281R)-mutant ZIKV (Extended Data Fig. 6e). Ubiquitinated E was also detected after immunoprecipitation of ZIKV grown in mosquito C6/36 cells (Extended Data Fig. 6e, lane 7), although it appeared more in shorter K63-linked polyubiquitin chains as compared to the longer ubiquitin chains found on E from wild-type ZIKV grown in JEG-3 cells (Extended Data Fig. 6e). DENV particles also contained ubiquitinated E (Extended Data Fig. 6f). Together, these data indicate that flaviviruses released from cells contain a proportion of ubiquitinated E.

In an effort to quantify the proportion of ubiquitinated ZIKV particles, we performed immunoprecipitations with an antibody against K63-linked ubiquitin and ZIKV supernatants obtained from different cells, and measured the proportion of viral RNA copies from input virus (Extended Data Fig. 7a). Approximately 5–6% of ZIKV particles grown in JEG-3 cells could be detected with the anti-K63-linked ubiquitin, which is significantly higher than the proportion of ubiquitinated E of ZIKV grown in Vero cells; ZIKV grown in JEG-3 TRIM7-knockout cells contained near to background levels of ubiquitinated E as compared to an IgG control (Extended Data Fig. 7a). Additional evidence that the intact virus particle contains a proportion of ubiquitinated E comes from cryo-electron microscopy studies. Ubiquitinated Zika virions were labelled using anti-K63-linked ubiquitin and a gold-anti-IgG secondary antibody. After purification of virus–antibody complexes by sucrose gradient, approximately 15% of virus particles showed at least 1 gold particle close to the virus (mostly between 50 and 150 nm away

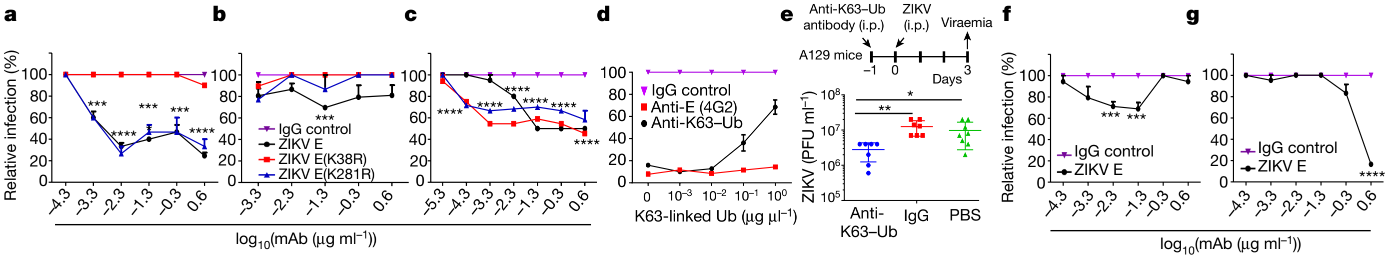


Fig. 4 | Specific anti-K63-linked polyubiquitin antibody neutralizes ZIKV in vitro and in vivo. a–**c**, **f**, **g**, Neutralization of ZIKV in Vero or C6/36 cells. Wild-type and mutant viruses grown in Vero cells (**a**–**c**) or mosquito C6/36 cells (**f**, **g**) were incubated at 37 °C with dilutions of anti-K63-linked ubiquitin (**a**, **f**), anti-K48-linked ubiquitin (**b**) or anti-E (4G2) (**c**, **g**) antibodies for 1 h, followed by a plaque assay. **d**, Competition assay. Purified K63-linked ubiquitin chains were incubated together with wild-type ZIKV and anti-K63-linked ubiquitin antibody, IgG control or 4G2. Relative infection was calculated as a percentage of antibody effects on each virus relative to its own IgG control. $n = 3$ technical replicates, mean ± s.e.m., two-way ANOVA, with Tukey correction. **e**, In vivo neutralization assay. A129 mice were inoculated intraperitoneally (i.p.) with anti-K63-linked ubiquitin or an IgG control. The next day mice were infected with ZIKV (1×10^4 PFU). Viraemia was assessed at day 3. $n = 7$, mean ± s.e.m. Unpaired two-sided *t*-test; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

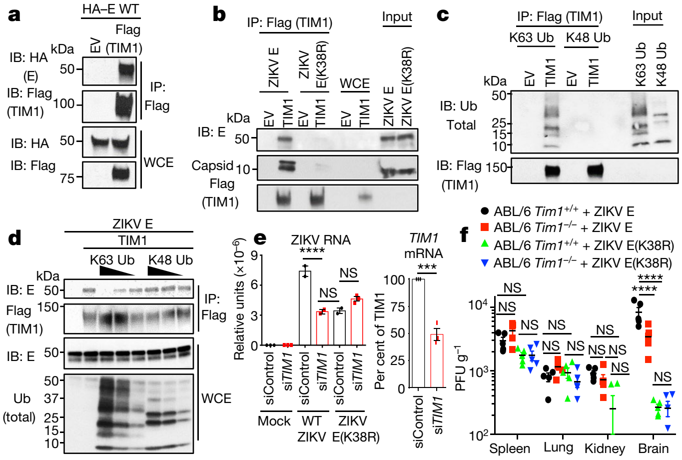


Fig. 5 | Ubiquitination of ZIKV E promotes binding to the TIM1 receptor. **a**, Whole-cell extract from HEK293T cells transfected with empty vector, HA–E and Flag–TIM1 were used for immunoprecipitation with anti-Flag beads. **b**, Interaction of wild-type ZIKV, but not E(K38R)-mutant ZIKV, viral particles with TIM1. Ectopically expressed TIM1 was isolated using anti-Flag beads. After washes, viruses were incubated with TIM1. **c**, Co-immunoprecipitation assay. Interaction of K63-linked, but not K48-linked, polyubiquitin with TIM1 in the absence of E. **d**, Wild-type ZIKV particles were bound to TIM1-containing beads (as in **b**). Increasing amounts of purified K63-linked ubiquitin or K48-linked ubiquitin chains were added. Representative of two independent experiments (**a**–**d**). **e**, Virus attachment (viral RNA, qPCR) after knockdown of *TIM1* (also known as *HAVCR1*) using siRNA in JEG-3 cells. $n = 2$, mean ± s.e.m., one-way ANOVA, multiple comparison, Tukey correction, $^{****}P < 0.0001$, $^{*}P < 0.05$, NS, not significant ($P > 0.05$)). siControl, non-targeting siRNA). **f**, Titres in infected *Tim1*$^{-/-}$ (also known as *Havcr1*) mice. $n = 5$ mice per group, mean ± s.e.m., two-way ANOVA, multiple comparison, Tukey correction. $^{****}P < 0.0001$, NS, not significant ($P > 0.05$). ABL/6, C57BL/6, *Ifnar1*$^{-/-}$.

from the viral particle) (Extended Data Fig. 7b). By contrast, no gold particles were found close to E(K38R/K281R)-mutant ZIKV under the same conditions.

TRIM7 determines ZIKV tissue tropism in vivo

To test the role of TRIM7 in ZIKV replication in vivo, we generated *Trim7*$^{-/-}$ mice (mutant sequences shown in Supplementary Fig. 2) and infections were performed using a previously established protocol[28] (Fig. 2g–i). Notably, viral titres in serum (Fig. 2h) and in kidney, eye, brain and reproductive tissues (uterus and testis) of *Trim7*$^{-/-}$ mice were significantly lower than those in wild-type mice (Fig. 2i). By contrast, ZIKV replicated to similar levels in tissues from wild-type and *Trim7*$^{-/-}$ mice, including heart, liver, lung and muscle (Fig. 2i), which indicates that TRIM7 promotes virus replication in a tissue-specific manner. These data suggest TRIM7 may be a determinant of tissue tropism.

E ubiquitination is important in virus entry

Because E mediates virus attachment to host cells and induces virus–endosome membrane fusion[15,29,30], we examined whether ubiquitination contributes to virus entry. Endosome–virus membrane fusion was analysed using a lipophilic dye (DiOC18) to label wild-type and mutant ZIKV. The ability of E(K38R)-mutant ZIKV to promote virus–endosome fusion was significantly decreased compared to that of wild-type ZIKV in both JEG-3 and A549 cells (Fig. 3a, Extended Data Figs. 8, 9). Ammonium chloride (NH_4Cl) treatment, which blocks acidification of the endosome and subsequent fusion, served as a control.

Although E(K281R)-mutant ZIKV did not significantly attenuate fusion in JEG-3 cells, it slightly reduced fusion in A549 cells (Extended Data Figs. 9b, c); this suggests that ubiquitination of E on K281 may affect virus entry and/or fusion in a cell-type-specific manner. Further evidence acquired by measuring the viral RNA of adsorbed viruses to cells indicates that the specific ubiquitination on K38 of E (and not on the K281 of E) is responsible for efficient virus attachment to the host cell (Fig. 3b; adsorption at 4 °C). Reduced levels of attachment of E(K38R)-mutant ZIKV, as compared to wild-type ZIKV, were also observed in human primary induced-pluripotent neural stem cells (Fig. 3c), brain microvascular endothelial cells and astrocytes (Extended Data Fig. 10a, b), which also correlated with reduced virus replication by plaque assay (Extended Data Fig. 10c–e). These effects are not due to reduced glycosylation of E because *N*-glycosidase F can cleave both ZIKV E(K38R) and E(K281R) mutants (Extended Data Fig. 10f). E(K38R)-mutant ZIKV grown in wild-type JEG-3 cells has reduced capacity to attach to cells as compared to wild-type ZIKV, and this is comparable to the reduced ability of ZIKV grown in TRIM7-knockout cells to attach to these cells (Extended Data Fig. 10g). To further rule out that the reduced attachment of E(K38R)-mutant ZIKV is due to any minor structural changes caused by the K-to-R mutation, and confirm that ubiquitination enhances entry and replication, we treated wild-type ZIKV with the deubiquitinase ovarian tumour (OTU) of Crimean-Congo haemorrhagic fever virus[31], which can cleave ubiquitin chains (Extended Data Fig. 10h). Ubiquitin removal with the OTU reduces ZIKV attachment as compared to an OTU (2A) mutant with reduced activity (Extended Data Fig. 10i). Treatment with this deubiquitinase also reduced the replication of wild-type ZIKV (Extended Data Fig. 10j).

ZIKV infection neutralized by anti-K63 antibody

We examined whether the monoclonal antibody against K63-linked polyubiquitin that we used in our immunoprecipitation experiments could neutralize ZIKV infection (Fig. 4a–c). Pretreatment with this antibody significantly decreased infection by wild-type and E(K281R)-mutant ZIKV—but not E(K38R)-mutant ZIKV-E—in a dose-dependent manner, as compared to an IgG control (Fig. 4a). This neutralizing effect was specific for K63-linked ubiquitin because antibodies against K48-linked polyubiquitin did not have major effects (Fig. 4b). As an additional control, we used a pan-flavivirus anti-E (4G2) monoclonal antibody, which neutralized wild-type ZIKV and ZIKV mutants—especially at higher antibody concentrations (Fig. 4c). The specificity of the anti-K63-linked polyubiquitin antibody in neutralizing ubiquitinated E was confirmed in a competition assay, in which addition of purified K63-linked polyubiquitin chains reduced the neutralizing activity of anti-K63-linked ubiquitin but not of anti-E (4G2) (Fig. 4d). Furthermore, administration of this anti-K63-linked ubiquitin antibody in mice one day before ZIKV infection significantly reduced virus titres, as compared to an IgG or PBS control, in vivo (Fig. 4e). Importantly, ZIKV produced in mosquito cells was less sensitive to neutralization by treatment with the anti-K63-linked ubiquitin antibody (Fig. 4f), whereas the anti-E 4G2 antibody inhibited this ZIKV at high concentrations (Fig. 4g).

E ubiquitination promotes binding to TIM1

Although their roles are unclear, multiple receptors—including DC-SIGN, AXL, TYRO3 and TIM1—have previously been proposed to mediate ZIKV attachment in specific cell types[32]. We tested whether ubiquitination of E may enhance affinity for TIM1. Co-immunoprecipitation assays revealed that recombinant TIM1 interacts with ectopically expressed ZIKV E (Fig. 5a) and infectious wild-type ZIKV viral particles, but only minimally with E(K38R)-mutant ZIKV viral particles (Fig. 5b); this suggests that ubiquitination on the K38 residue is responsible for the interaction. In support of this, recombinant purified K63-linked, but not K48-linked, polyubiquitin chains interact with TIM1 in the absence of E (Fig. 5c). Furthermore, K63-linked polyubiquitin chains compete with ZIKV particles for interaction with TIM1, whereas K48-linked polyubiquitin chains do not (Fig. 5d). Ubiquitinated E is also likely to mediate virus attachment to cells—at least in part via TIM1—because knockdown of TIM1 in JEG-3 cells significantly reduced the levels of attachment of wild-type ZIKV, whereas the attachment of E(K38R)-mutant ZIKV was not further reduced in TIM1-knockdown cells as compared to control cells (Fig. 5e). Finally, infection of *Havcr1*$^{-/-}$ (the gene that encodes the TIM1 protein) mice (C57BL/6, *Ifnar1*$^{-/-}$ background) with wild-type ZIKV exhibited a small (approximately 2.5-fold)—but significant—reduction in virus titres in the brain as compared to *Havcr1*$^{+/+}$ controls (Fig. 5f). By contrast, although replication of E(K38R)-mutant ZIKV was strongly reduced as compared to wild-type ZIKV in the brain, no difference was observed between wild-type and *Havcr1*$^{-/-}$ mice. In addition, no differences were observed in other tissues (for example, spleen, lung or kidney) between wild-type and E(K38R)-mutant ZIKV and between *Havcr1*$^{+/+}$ and *Havcr1*$^{-/-}$ mice (Fig. 5f). The data suggest that although TIM1 is not the only receptor that mediates entry of ZIKV, it may have a role in specific cell types or tissues (such as the brain). Taken together, K63-linked ubiquitination on the K38 residue of E promotes efficient virus attachment to host receptors, and it is at least in part mediated by TIM1.

Discussion

We have shown that ubiquitinated E present in infectious virions of ZIKV promotes efficient entry of the virus into host cells; however, this ubiquitination is not a requirement for virus replication, as both of the ZIKV mutants we tested were attenuated but still able to replicate. Our data support a model in which ubiquitination on the K38 residue of E enhances viral attachment to host cell receptors, thereby increasing the efficiency of virus replication. This occurs in a tissue-specific manner, and could partially be explained by the expression levels of TRIM7 in combination with other factors such as the expression of the E2 conjugase UBCH5A or additional cellular receptors—which may also contribute to the characteristic ZIKV tropism. Our data show ubiquitination on residue K38 (which is conserved among members of the *Flaviviridae*[33]), and—combined with the fact that DENV particles also contain K63-linked polyubiquitinated E—this raises the possibility that ubiquitination on K38 may be used as a general mechanism in the entry of flaviviruses into host cells. We also identified an additional ubiquitination site on K281 of the E of ZIKV that, to our knowledge, is not present in other flaviviruses. The combined ubiquitination on both residues could contribute to the differential tropism between ZIKV and other flaviviruses.

Despite existing literature on the crucial role of E in binding to host-cell receptors or neutralizing antibodies, and structural studies that include ZIKV or E from other flaviviruses[14,34,35], to our knowledge no previous studies have detected ubiquitination on E of flaviviruses. One possible explanation is that many structural studies have used ZIKV prepared from mosquito cells, which—according to our data—may produce virions with reduced levels of ubiquitination. The presence of only a small proportion of ubiquitinated E on the viral particle, as suggested by our cryo-electron microscopy data, may also contribute to previous observations of imperfect virion symmetry[36]. Finally, our studies indicate that the anti-K63-linked ubiquitin antibody has neutralizing activity in vivo and could provide a therapeutic approach against ZIKV.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2457-8.

1. Musso, D. et al. Potential sexual transmission of Zika virus. *Emerg. Infect. Dis*. **21**, 359–361 (2015).
2. Hills, S. L. et al. Transmission of Zika virus through sexual contact with travelers to areas of ongoing transmission — continental United States, 2016. *MMWR Morb. Mortal Wkly Rep*. **65**, 215–216 (2016).
3. Driggers, R. W. et al. Zika virus infection with prolonged maternal viremia and fetal brain abnormalities. *N. Engl. J. Med*. **374**, 2142–2151 (2016).
4. van Tol, S., Hage, A., Giraldo, M. I., Bharaj, P. & Rajsbaum, R. The TRIMendous role of TRIMs in virus–host interactions. *Vaccines (Basel)* **5**, 23 (2017).
5. Byk, L. A. et al. Dengue virus genome uncoating requires ubiquitination. *MBio* **7**, e00804-16 (2016).
6. Fernandez-Garcia, M. D. et al. Appraising the roles of CBLL1 and the ubiquitin/proteasome system for flavivirus entry and replication. *J. Virol*. **85**, 2980–2989 (2011).
7. Choy, M. M. et al. Proteasome inhibition suppresses dengue virus egress in antibody dependent infection. *PLoS Negl. Trop. Dis*. **9**, e0004058 (2015).
8. Versteeg, G. A. et al. The E3-ligase TRIM family of proteins regulates signaling pathways triggered by innate immune pattern-recognition receptors. *Immunity* **38**, 384–398 (2013).
9. Hage, A. & Rajsbaum, R. To TRIM or not to TRIM: the balance of host–virus interactions mediated by the ubiquitin system. *J. Gen. Virol*. **100**, 1641–1662 (2019).
10. Bharaj, P. et al. The host E3-ubiquitin ligase TRIM6 ubiquitinates the Ebola Virus VP35 protein and promotes virus replication. *J. Virol*. **91**, e00833-17 (2017).
11. Fink, J. et al. Host gene expression profiling of dengue virus infection in cell lines and patients. *PLoS Negl. Trop. Dis*. **1**, e86 (2007).
12. Padilla-S, L., Rodríguez, A., Gonzales, M. M., Gallego-G, J. C. & Castaño-O, J. C. Inhibitory effects of curcumin on dengue virus type 2-infected cells in vitro. *Arch. Virol*. **159**, 573–579 (2014).
13. Choy, M. M., Sessions, O. M., Gubler, D. J. & Ooi, E. E. Production of infectious dengue virus in *Aedes aegypti* is dependent on the ubiquitin proteasome pathway. *PLoS Negl. Trop. Dis*. **9**, e0004227 (2015).
14. Kostyuchenko, V. A. et al. Structure of the thermally stable Zika virus. *Nature* **533**, 425–428 (2016).
15. Pierson, T. C. & Kielian, M. Flaviviruses: braking the entering. *Curr. Opin. Virol*. **3**, 3–12 (2013).
16. Rossi, S. L. et al. Characterization of a novel murine model to study Zika virus. *Am. J. Trop. Med. Hyg*. **94**, 1362–1369 (2016).
17. Zhao, Z. et al. Viral retinopathy in experimental models of Zika infection. *Invest. Ophthalmol. Vis. Sci*. **58**, 4355–4365 (2017).
18. Montori-Grau, M. et al. GNIP1 E3 ubiquitin ligase is a novel player in regulating glycogen metabolism in skeletal muscle. *Metabolism* **83**, 177–187 (2018).
19. Le Sommer, C., Barrows, N. J., Bradrick, S. S., Pearson, J. L. & Garcia-Blanco, M. A. G protein-coupled receptor kinase 2 promotes flaviviridae entry and replication. *PLoS Negl. Trop. Dis*. **6**, e1820 (2012).
20. Skurat, A. V., Dietrich, A. D., Zhai, L. & Roach, P. J. GNIP, a novel protein that binds and activates glycogenin, the self-glucosylating initiator of glycogen biosynthesis. *J. Biol. Chem*. **277**, 19331–19338 (2002).
21. Zhai, L., Dietrich, A., Skurat, A. V. & Roach, P. J. Structure–function analysis of GNIP, the glycogenin-interacting protein. *Arch. Biochem. Biophys*. **421**, 236–242 (2004).
22. Orchard, R. C. et al. Identification of antinorovirus genes in human cells using genome-wide CRISPR activation screening. *J. Virol*. **93**, e01324-18 (2018).
23. Lu, M. et al. E3 ubiquitin ligase tripartite motif 7 positively regulates the TLR4-mediated immune response via its E3 ligase domain in macrophages. *Mol. Immunol*. **109**, 126–133 (2019).
24. Chakraborty, A., Diefenbacher, M. E., Mylona, A., Kassel, O. & Behrens, A. The E3 ubiquitin ligase Trim7 mediates c-Jun/AP-1 activation by Ras signalling. *Nat. Commun*. **6**, 6782 (2015).
25. Napolitano, L. M., Jaffray, E. G., Hay, R. T. & Meroni, G. Functional interactions between ubiquitin E2 enzymes and TRIM proteins. *Biochem. J*. **434**, 309–319 (2011).
26. Chazotte, B. Labeling membrane glycoproteins or glycolipids with fluorescent wheat germ agglutinin. *Cold Spring Harb. Protoc*. **2011**, pdb.prot5623 (2011).
27. Rossignol, E. D., Peters, K. N., Connor, J. H. & Bullitt, E. Zika virus induced cellular remodelling. *Cell. Microbiol*. **19**, e12740 (2017).
28. Gorman, M. J. et al. An immunocompetent mouse model of Zika virus infection. *Cell Host Microbe* **23**, 672–685 (2018).
29. Fontes-Garfias, C. R. et al. Functional analysis of glycosylation of Zika virus envelope protein. *Cell Rep*. **21**, 1180–1190 (2017).
30. Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427**, 313–319 (2004).
31. Frias-Staheli, N. et al. Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses. *Cell Host Microbe* **2**, 404–416 (2007).
32. Hamel, R. et al. Biology of Zika virus infection in human skin cells. *J. Virol*. **89**, 8880–8896 (2015).
33. Dai, L. et al. Structures of the Zika virus envelope protein and its complex with a flavivirus broadly protective antibody. *Cell Host Microbe* **19**, 696–704 (2016).
34. Sirohi, D. et al. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* **352**, 467–470 (2016).
35. Fibriansah, G. et al. Cryo-EM structure of an antibody that neutralizes dengue virus type 2 by locking E protein dimers. *Science* **349**, 88–91 (2015).
36. Therkelsen, M. D. et al. Flaviviruses have imperfect icosahedral symmetry. *Proc. Natl Acad. Sci. USA* **115**, 11608–11612 (2018).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All data are available in the Article and its Supplementary Information. Mutant viruses may be available upon request to the corresponding authors, after respective material transfer agreements are completed. Source data are provided with this paper.

Acknowledgements We thank M. A. Garcia-Blanco, S. Bradrick and R. Soto for their generosity in sharing reagents and advice; Y. Liang for his technical advice on flow cytometry; M. Fan for providing the lentivirus to establish stable cell lines that express HA–Ub (pLenti puro HA-Ub), through Addgene; A. Gamarnik, R. Stephens and V. Menachery for suggestions and helpful discussions; T. Wang for providing some control mice; and L. Yeager for editing. The laboratory of R.R. was supported in part by the John Sealy Memorial Endowment Fund for Biomedical Research (UTMB), a research career development award (K12HD052023: BIRCWH program, from NIH ORWH/NICHD), and NIH/NIAID grants R21 AI132479-01, R21 AI126012-01A1 and R01 AI134907-01. S.V.T. was supported by NIH/NIAID T32-AI060549, and A.H. by NIH/NIAID T32 AI007526. P.-Y.S. was supported by NIH grants AI142759, AI134907, AI145617 and UL1TR001439, and awards from the Sealy & Smith Foundation, Kleberg Foundation, John S. Dunn Foundation, Amon G. Carter Foundation, Gilson Longenbaugh Foundation and Summerfield Robert Foundation. The laboratory of S.M.B. was supported in part by the Division of Intramural Research of the NIH/NIAID. J.R.J. and N.J.K. were supported by NIH/NIAID grant U19 AI118610.

Author contributions M.I.G. performed all aspects of this study. M.I.G., H.X., L.A.-A., S.R.A., S.L.R., A.H., S.v.T., C.S. and X.X. performed experiments and analysed data. H.X. generated ZIKV recombinant viruses. S.R.A. and S.L.R. performed in vivo experiments. P.-Y.S. and R.R. organized, conceptualized the study and provided funding. G.L.S., S.J.R., K.L.M., K.M.-W. and S.M.B. generated the *Trim7*$^{-/-}$ mice. W.M. provided the AB6 *Tim-1*$^{-/-}$ mice. M.W. and M.C.M. performed cryo-electron microscopy. H.R., J.R.J. and N.J.K. performed mass spectrometry studies. M.I.G., L.A.-A, P.-Y.S. and R.R. prepared the manuscript. All authors discussed the results and commented on the manuscript.

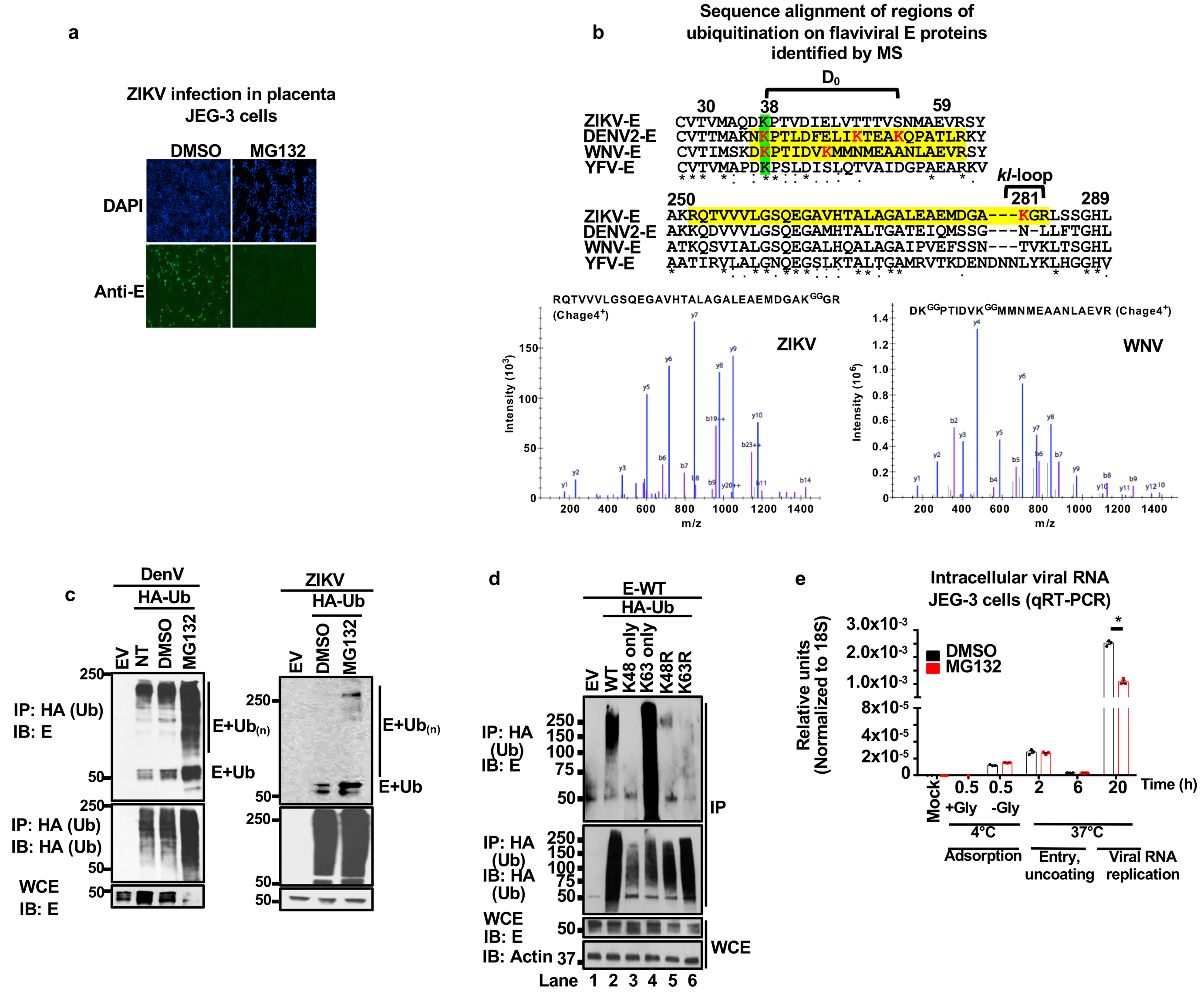
Competing interests The authors declare no competing interests.

Additional information

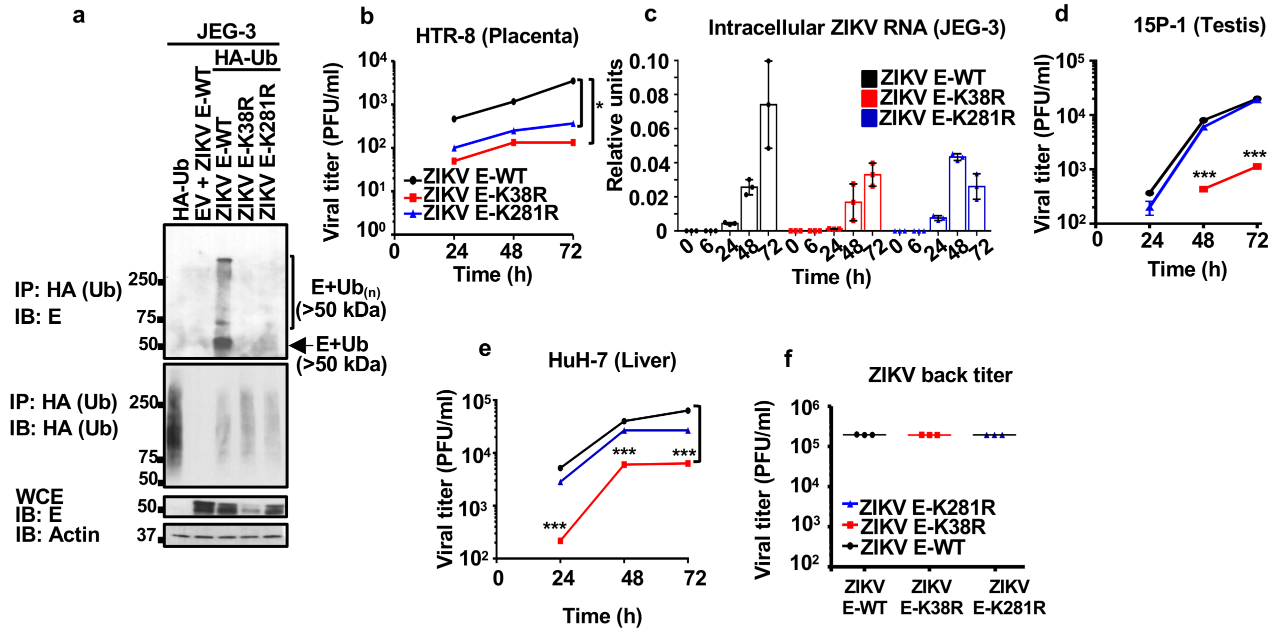
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2457-8.

Correspondence and requests for materials should be addressed to P.-Y.S. or R.R.

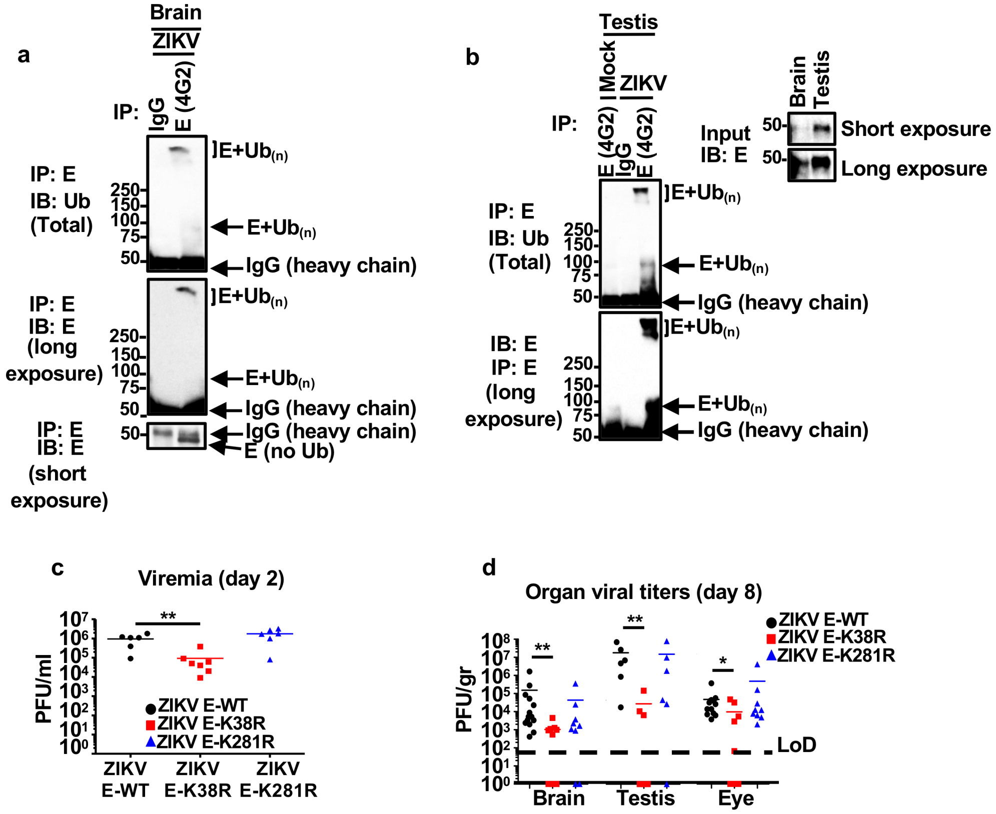
Peer review information *Nature* thanks Jan Carette and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



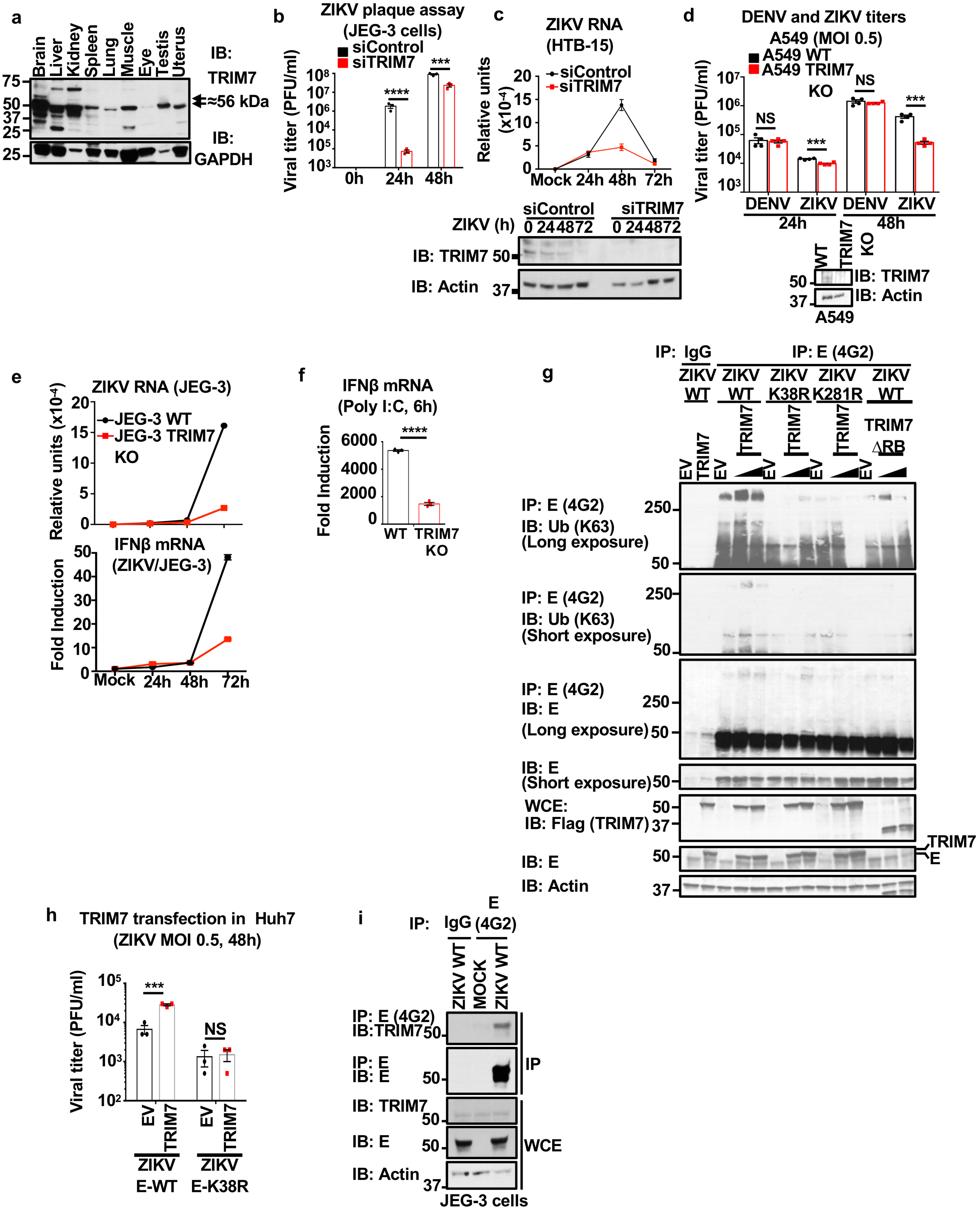
Extended Data Fig. 1 | Ubiquitination of flavivirus E protein. **a**, Proteasome inhibition blocks ZIKV replication. JEG-3 cells were pretreated with DMSO or MG132 (2 h) followed by ZIKV infection (MOI 2, 24 h, visualized by immunofluorescence with anti-E 4G2). **b**, Ubiquitinated peptides from flavivirus-infected cells identified by mass spectrometry (peptides highlighted in yellow, diglycine residues indicating ubiquitination in red and conserved residues in green). Sequences for strains ZIKV FSS13025, GenBank: KU955593.1; DENV-2 Y98P, JF327392.1; West Nile virus (WNV) NY99, DQ211652; and yellow fever virus (YFV), ANC33490.1. JEG-3 cells were used for ZIKV infections, Huh7 cells for DENV infections and A549 cells for WNV infections (repeated in U2OS cells with identical results, two independent experiments). Representative mass spectra for ubiquitinated peptides found for WNV are shown. *b* and *y* ions are indicated in blue and red, respectively. **c**, Whole-cell extracts from DENV- or ZIKV-infected (MOI 2, 20 h) Huh7 cells transfected with HA–Ub, followed by DMSO or MG132 treatment (6 h) were used for HA immunoprecipitation. Immunoblots are shown. NT, non-treated **d**, Whole-cell extract from cells transfected with vectors expressing E and wild-type ubiquitin or all K-to-R mutants except for K63 or K48 (only), or Ub(K48R) and Ub(K63R), followed by immunoprecipitation. **e**, JEG-3 cells pretreated with MG132 or DMSO. Cells were incubated with ZIKV at 4 °C for 30 min, followed by a wash with or without glycine to test virus adsorption. Additional samples were then switched to 37 °C to allow virus internalization. Viral RNA detection by quantitative reverse-transcription PCR (qRT–PCR). Representative of two independent experiments, $n = 3$ technical replicates, mean ± s.e.m., unpaired *t*-test, two-sided, $^*P < 0.05$; NS, not significant. All experiments are representatives from two independent experiments, with similar results.



Extended Data Fig. 2 | Differences in Zika virus cell tropism are associated with ubiquitination of E. a, JEG-3 cells stably expressing HA–Ub were infected with wild-type ZIKV, or recombinant infectious E(K38R)- and E(K281R)-mutant ZIKV. Whole-cell extracts were used for immunoprecipitation with HA beads (same experiment as shown in Fig. 1b, but without normalizing the input for the immunoprecipitation). Reduced replication of E(K38R)-mutant ZIKV can be seen represented by the levels of E in the whole-cell extract. Representative of two independent experiments. **b**–**e**, Different cell types were infected with either wild-type ZIKV, E(K38R)- or E(K281R)-mutant ZIKV (MOI 0.5). Cells were lysed for RNA extraction and virus quantification by qRT–PCR (**c**), and supernatants were collected for plaque assays at different time-points, for HTR-8 (**b**), 15P-1 (**d**) and HuH7 (**e**) cells. **f**, Back titration for the virus used on these experiments, and for Fig. 1e, f. Representatives from two independent experiments. $n = 3$ technical replicates, mean ± s.e.m., multiple *t*-test, Holm–Sidak correction, $^{*}P < 0.05$, $^{***}P < 0.001$.

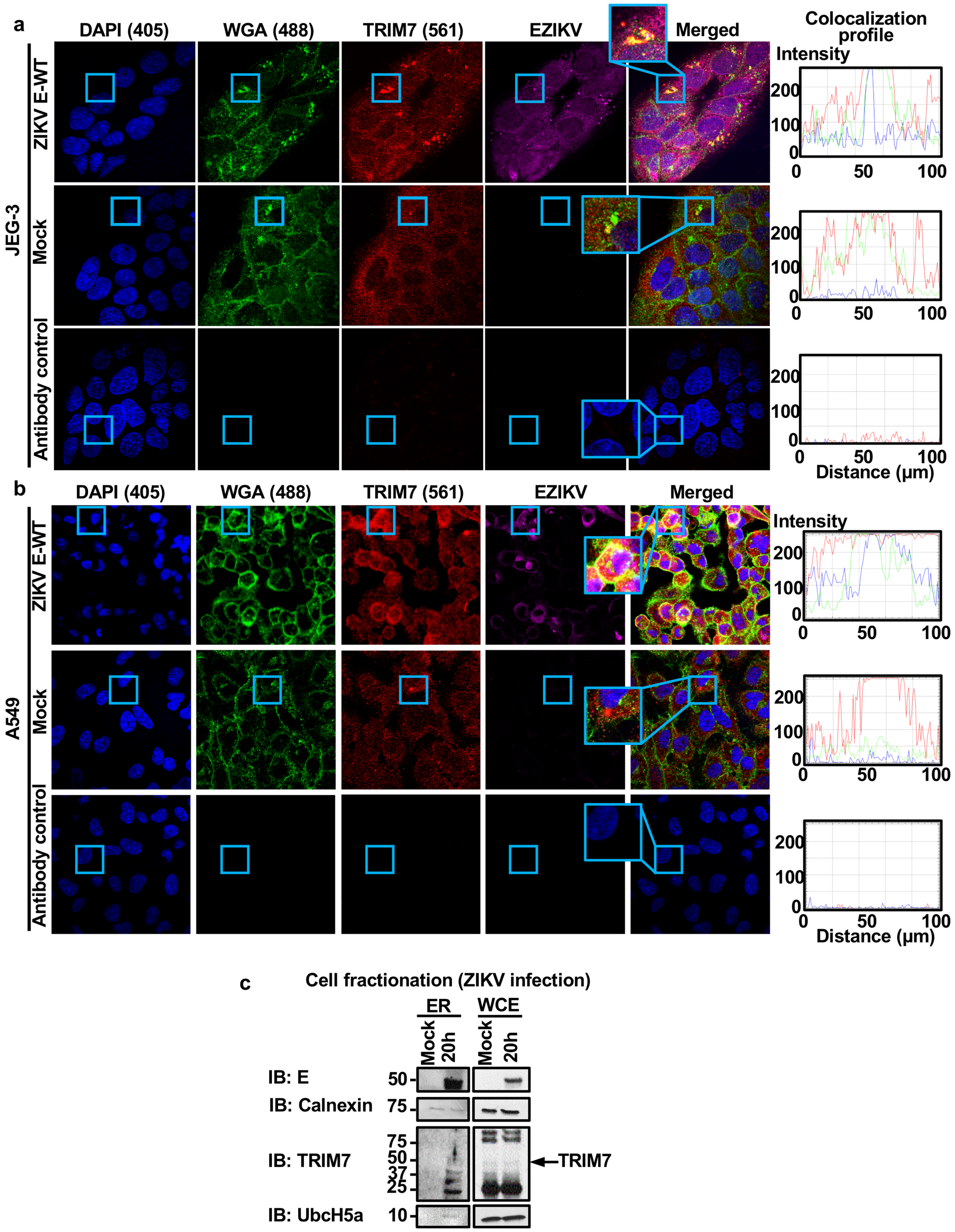


Extended Data Fig. 3 | Ubiquitination of E in tissues from infected mice. **a**, **b**, Tissues from testis (**a**) and brain (**b**) from mock-infected or wild-type-ZIKV-infected A129 mice were collected at day 8 after infection. Tissues were homogenized and 200 µg of total input protein was used for immunoprecipitation of E using 4G2 antibody or an IgG control. Ubiquitination of wild-type E was detected with anti-ubiquitin antibody by immunoblot. Immunoprecipitations shown are from mixed tissue lysates from three different mice. **c**, **d**, A129 mice (male and females) were mock-treated (5 mice) or infected with wild-type ZIKV, E(K38R)- or E(K281R)-mutant ZIKV (1×10^4 PFU, 9 mice per group, combined from 2 independent experiments). Weight loss and survival is shown in Fig. 1e, f. **c**, Serum titres (viraemia), were determined at day 2 after infection by plaque assay, after blood collection from 6 mice for wild-type and E(K281R)-mutant ZIKV, and 7 mice for E(K38R)-mutant ZIKV. **d**, Virus titres (at day 8 after infection) in brain (14 mice for wild-type ZIKV, and 9 mice for E(K38R)- and E(K281R)-mutant ZIKV), testis (6 mice per group) and eye (14 mice for wild-type ZIKV, and 9 mice for E(K38R)- and E(K281R)- ZIKV). Unpaired, *t*-test, two-sided, $*P < 0.05$, $**P < 0.01$.

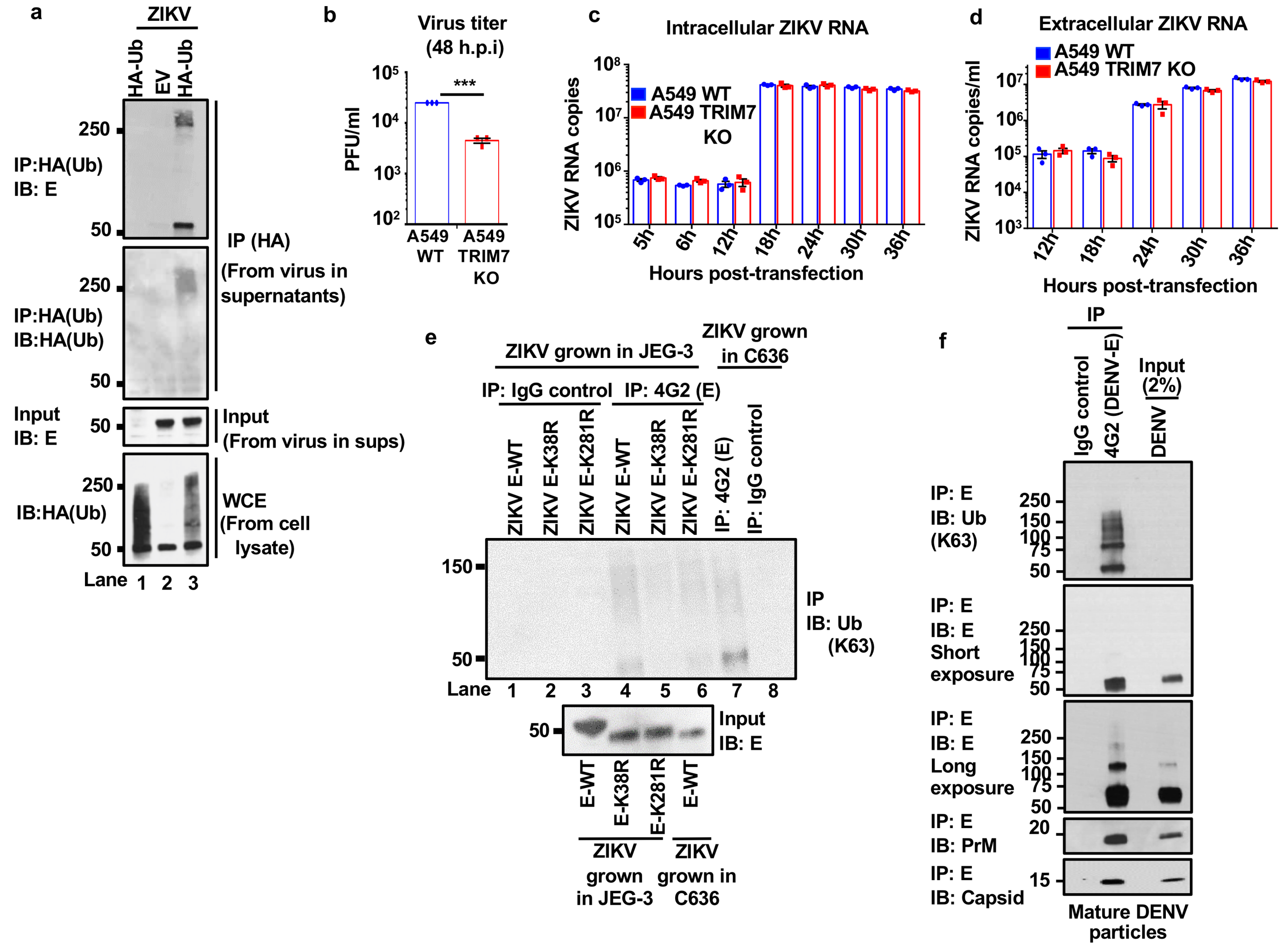


Extended Data Fig. 4 | See next page for caption.

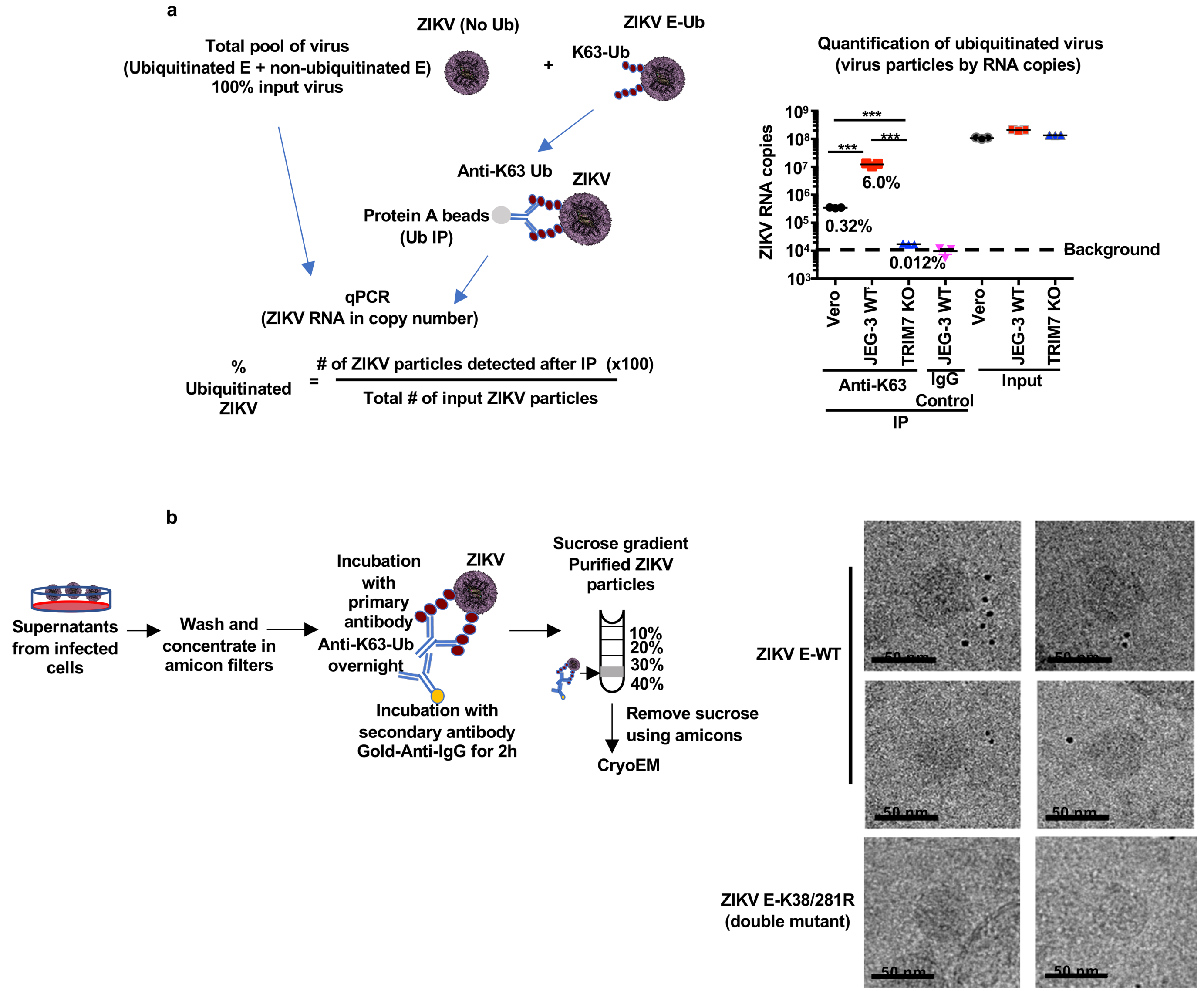
Extended Data Fig. 4 | TRIM7 interacts with and ubiquitinates E and promotes virus replication. a, Differential expression of TRIM7 in mouse tissues by immunoblot. The predicted molecular weight of full-length TRIM7 is 56 kDa. **b**, **c**, TRIM7 knockdown (24 h) in JEG-3 (**b**) or HTB-15 (**c**) cells followed by infection with ZIKV (MOI 1). **c**, Viral RNA levels were determined by qRT–PCR at different time points (top). TRIM7 knockdown efficiency was confirmed by western blot (bottom). **d**, TRIM7-knockout A549 and wild-type parental cells were used for infections with ZIKV or DENV at an MOI of 0.5. Bottom, immunoblot of TRIM7. Plaque assays from supernatants collected at different time points are shown. **e**, **f**, Infections of wild-type and TRIM7-knockout JEG-3 cells with ZIKV (MOI 0.5) or poly(I:C) stimulation (**f**, transfection of 10 μg ml^{-1} with Lipofectamine 2000). Quantification of ZIKV RNA (**e**, top) and *IFNB1* mRNA expression (**e**, bottom and **f**) by qPCR. **g**, Overexpression of TRIM7 enhances K63-linked polyubiquitination of wild-type E but not E(K38R) or E(K281R). HEK293T cells were transfected with vectors expressing wild-type E, E(K38R) or E(K281R) and different amounts of HA–TRIM7 (350 ng or 700 ng). Thirty hours after transfection, cells lysates were used for immunoprecipitation with anti-E 4G2 or isotype control. Immunoblots with indicated antibodies. **h**, Transfection of Huh7 cells with empty vector or vector expressing TRIM7. After 48 h cells were infected with wild-type or E(K38R)-mutant ZIKV. $n = 3$ technical replicates, mean ± s.e.m. Multiple *t*-test, two-sided, ***$P < 0.001$, ****$P < 0.0001$, NS, not significant ($P > 0.05$). **i**, Endogenous TRIM7 interacts with E in ZIKV-infected JEG-3 cells. Cells were infected with wild-type ZIKV (MOI 2). Thirty hours after infection, cells were lysed and whole-cell extracts were used for immunoprecipitation with anti-E (4G2) or isotype control. Representative of two independent experiments.



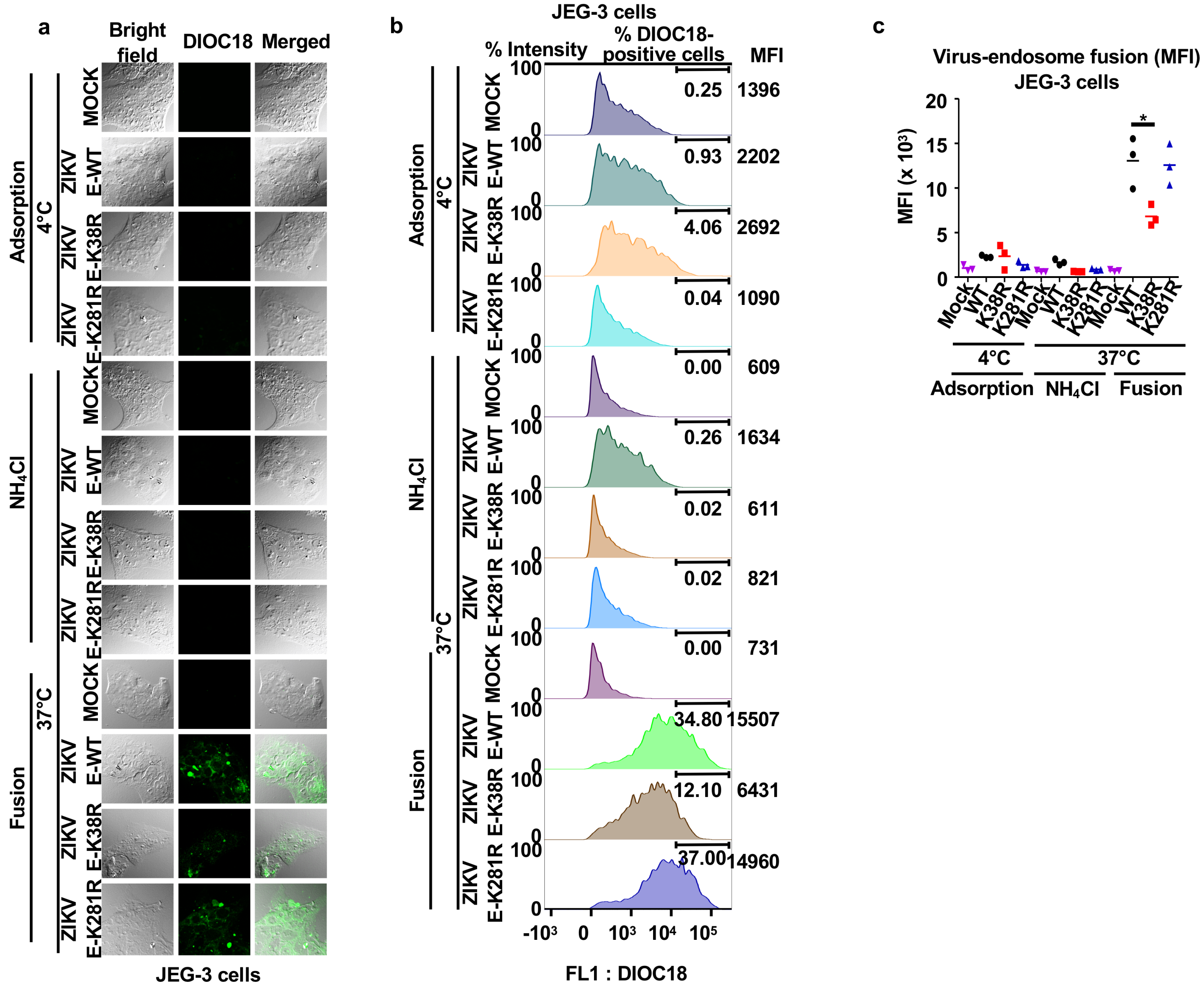
Extended Data Fig. 5 | TRIM7 colocalizes with E in the Golgi. a, **b**, JEG-3 (**a**) or A549 (**b**) cells were mock-treated or infected with ZIKV (MOI 2). Twenty-four hours after infection, cells were fixed and stained for endogenous TRIM7 (red), Golgi (WGA–FITC, green) and E (4G2, purple) for confocal microscopy. Colocalization is shown in rectangles, and red–green–blue (RGB) profile graphs are on the right. All images were processed identically using the same conditions with ZEN 2.5.75.0 (Zeiss), and RGB profiles were obtained using ImageJ v1.52e (NIH). **c**, Cell fractionation of infected JEG-3 cells (20 h, MOI 2) for endoplasmic reticulum (ER) was performed following the manufacturer's instructions (Sigma). Representative of two independent experiments.



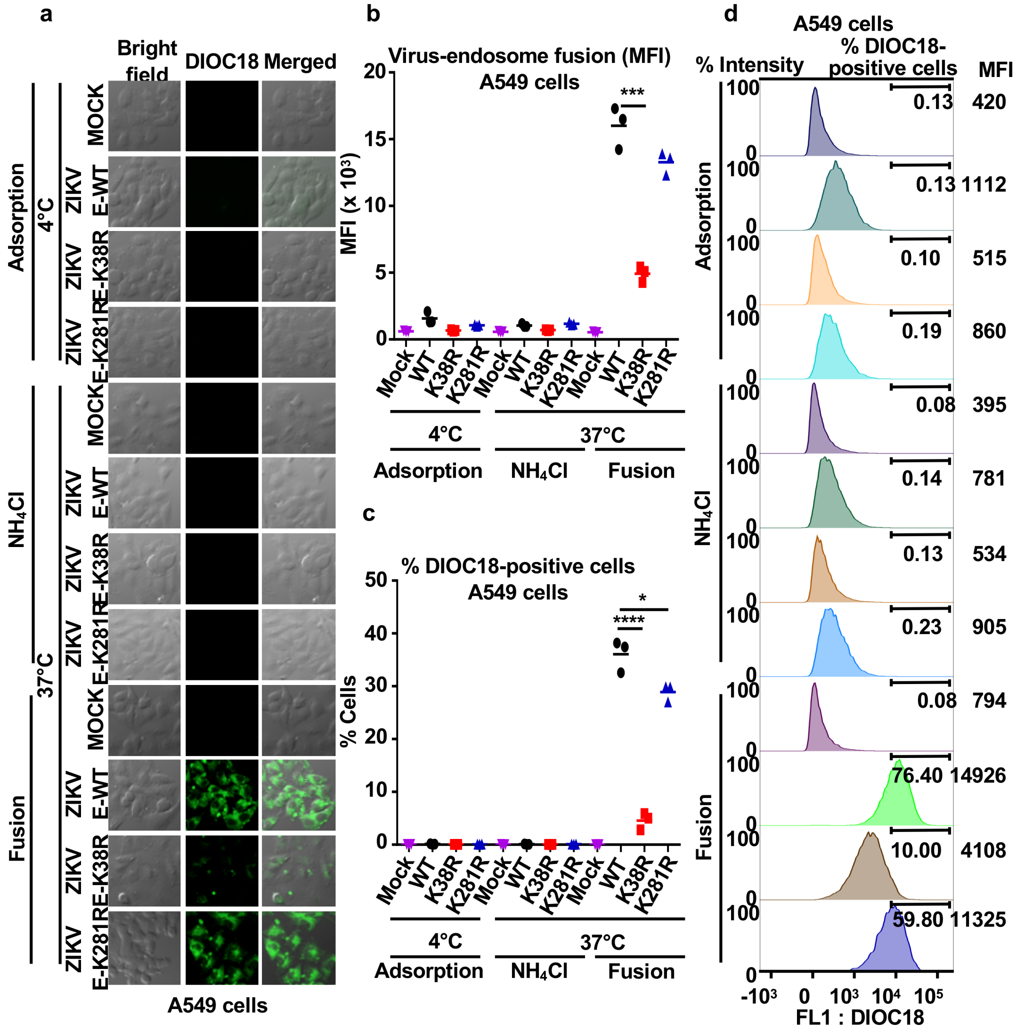
Extended Data Fig. 6 | Virions of ZIKV and DENV contain ubiquitinated E, and TRIM7 is important for ZIKV—but not DENV—replication. **a**, Supernatants collected from JEG-3 ZIKV-infected cells stably expressing HA–Ub showed detectable levels of ubiquitinated E. Viruses were immunoprecipitated with anti-HA beads. Immunoblot for E and HA–Ub. The bottom panel shows the total HA–Ub levels expressed in these HA–Ub stable cell lines. **b**, Infection of wild-type A549 or TRIM7-knockout cells with ZIKV. **c**, **d**, Transfection of wild-type A549 or TRIM7-knockout cells with in vitro-transcribed ZIKV RNA. Viral RNA was quantified by qPCR from cell lysates (intracellular) (**c**) or from supernatants (extracellular) (**d**). Representative of two independent experiments. *n* = 3 technical replicates, mean ± s.e.m. **e**, K63-linked polyubiquitinated E was detected on wild-type ZIKV particles but reduced in E(K38R)- and E(K281R)-mutant ZIKV particles, after immunoprecipitation with anti-E 4G2 antibody. Although reduced, ubiquitination on E from ZIKV grown in mosquito C6/36 cells can also be detected. **f**, DENV from supernatants from infected BHK-21 cells was immunoprecipitated with anti-E antibody (4G2) or an IgG control. Immunoblot for K63-linked polyubiquitin and each viral protein are shown. Representative of two independent experiments.



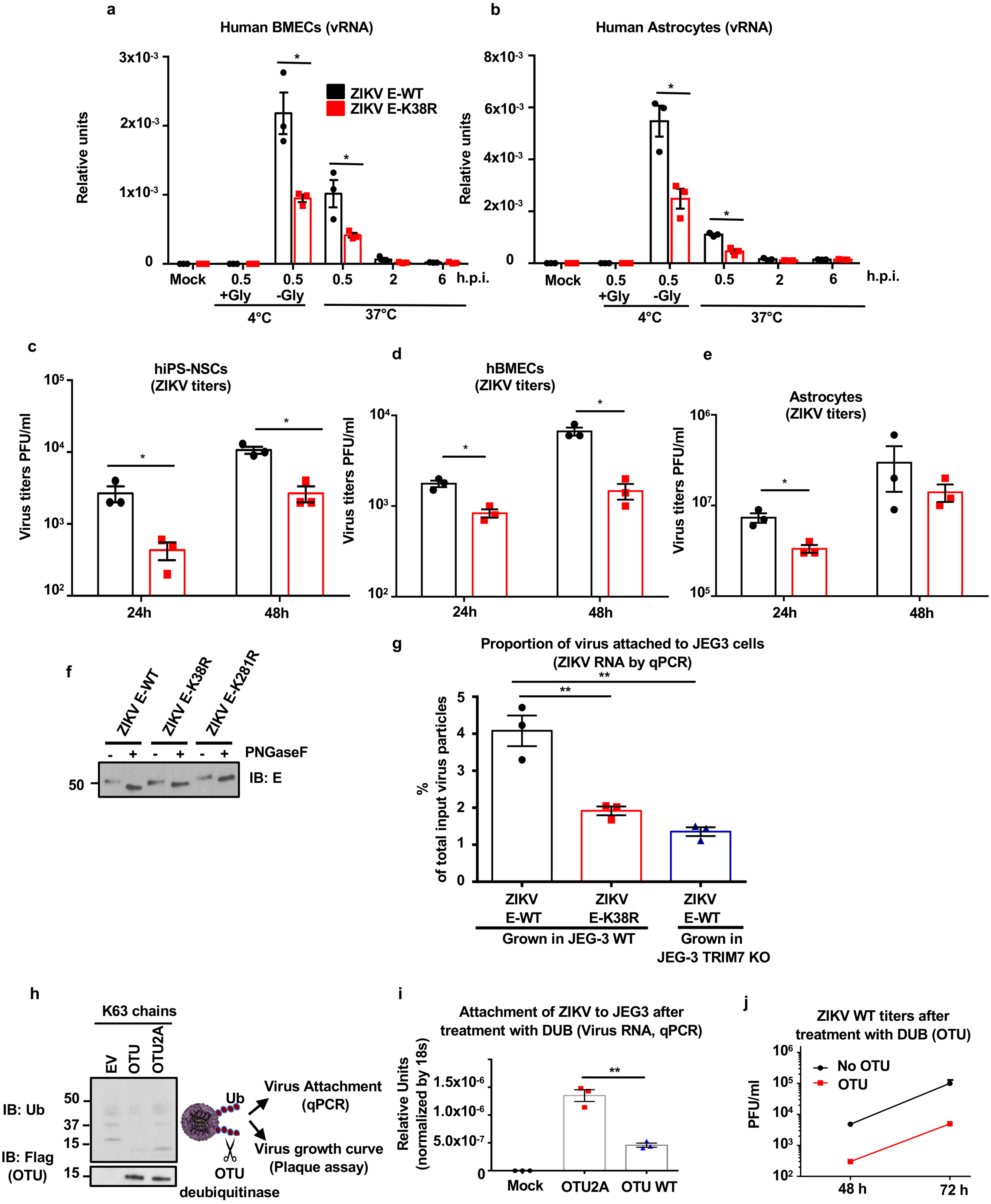
Extended Data Fig. 7 | Proportion of ubiquitinated E in virions of ZIKV, and cryo-electron microscopy of ubiquitinated ZIKV. a, **b**, ZIKV stocks were grown in Vero cells, wild-type JEG-3 or TRIM7-knockout JEG-3 cells, and used for immunoprecipitation using an anti-K63-linked ubiquitin antibody, or an IgG control to set the background levels. The immunoprecipitated virus, as well as a sample of input viruses, was lysed in Trizol for virus RNA quantification by qPCR (**a**). The virus RNA copy number was determined using a standard of purified ZIKV RNA and its known molecular weight. The proportion of ubiquitinated virus was calculated taking as 100% the input virus. $n = 3$ technical replicates, mean, unpaired two-sided *t*-test, $^{***}P < 0.001$. **b**, Cryo-electron microscopy of ubiquitinated ZIKV. Experimental approach. Supernatants from Vero cells infected with wild-type or E(K38R/K281R)-mutant ZIKV were washed and concentrated in Amicon filters followed by labelling with a primary antibody against K63-linked ubiquitin and secondary nano-gold-labelled antibody. Virus–antibody complexes were then purified by sucrose gradient. A visible band containing these complexes was recovered and passed through Amicon filters to remove sucrose, and concentrate the complexes. Samples were flash-frozen in liquid ethane cooled to liquid nitrogen temperatures on holey carbon grids and images were recorded in movie mode at 40,000× magnification using a 200 KV JEOL 2200FS transmission electron microscope. To facilitate visualization of virus particles, frames were further binned 3× to yield a pixel size of 4.398 Å per pixel. These binned micrographs were manually examined using EMAN2. To identify potentially gold-labelled ubiquitinated particles, we looked for spherical particles corresponding to the known approximately 500 Å (50 nm) size of mature ZIKV, and which were within 200 Å of the easily recognizable nano-gold clusters. Approximately 15% of visible approximately 500 Å wild-type ZIKV particles satisfied these criteria. None of the E(K38R/K281R)-mutant ZIKV was found labelled with gold particles (**b**). The cryo-electron microscopy experiments with gold particle labelling were performed only once, owing to the large amount of virus needed.



Extended Data Fig. 8 | Ubiquitination of E promotes virus–endosome membrane fusion. These experiments were performed in JEG-3 cells. **a**–**c**, Wild-type ZIKV, E(K38R)- and E(K281R)-mutant ZIKV were labelled with DiOC18. After filtration, viruses were incubated at 4 °C with JEG-3 cells at MOI 2, and after 30 min were washed and collected for controls as adsorbed viruses. Additional samples were then incubated at 37 °C for 1 h, in the presence or absence of NH_4Cl to block acidification (as control), washed, fixed and visualized with a confocal microscope (**a**). The same experiment was repeated for quantification by fluorescence-activated cell sorting (FACS). Mean fluorescence intensity (MFI) is shown in **b**; the percentage of cells infected is shown in main Fig. 3a; and representative histograms are shown in **c**. $n = 3$ technical replicates, mean, unpaired two-sided t-test, $^*P < 0.05$. Representative of two independent experiments.



Extended Data Fig. 9 | Ubiquitination of E promotes virus–endosome membrane fusion in A549 cells. a–**d**, Wild-type ZIKV, E(K38R)- and E(K281R)-mutant ZIKV were labelled with DiOC18. After filtration, viruses were incubated at 4 °C with A549 cells at MOI 2 and after 30 min were washed and collected for controls as adsorbed viruses. Additional samples were then incubated at 37 °C for 1 h, and NH_4Cl was used as control to block acidification of the endosome. Cells were washed, fixed and visualized in a confocal microscope in **a**. Quantification by FACS is shown (mean fluorescence intensity) in **b**, **c**. **d**, Representative histograms showing green fluorescence (FL1: DiOC18) during ZIKV–endosome fusion. *n* = 3 technical replicates, mean, unpaired two-sided *t*-test, $^*P < 0.05$, $^{***}P < 0.05$. Representative of two independent experiments.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Ubiquitination of E on K38 promotes ZIKV attachment and enhanced replication in relevant human cells. a–**e**, Human brain microvascular endothelial cells (BMECs) (**a**, **d**), human astrocytes (**b**, **e**) and human primary induced-pluripotent neural stem cells (hiPS-NSCs) (**c**) were infected with wild-type or E(K38R)-mutant ZIKV (MOI 2), as described in Fig. 3b, c. Viral RNA was quantified by qPCR (**a**, **b**) and virus titres by plaque assay (**c**–**e**). One experiment, $n = 3$ technical replicates, mean ± s.e.m. Unpaired two-sided t-test, $^{*}P < 0.05$. **f**, Endoglycosidase analyses of E. Proteins from wild-type ZIKV and E(K38R)- and E(K281R)-mutant ZIKV were analysed by western blot. Viruses were treated with PNGase F for 1 h at 37 °C. **g**, Wild-type ZIKV or E(K38R)-mutant ZIKV grown in wild-type or TRIM7-knockout JEG-3 cells were used for attachment assays. Viruses were incubated at 4 °C for 30 min with JEG-3 cells and attachment was determined by measuring virus RNA by qPCR. The percentage of virus attachment was calculated by taking the input virus as 100%. **h**–**j**, The deubiquitinase (DUB) domain of the OTU of the Crimean-Congo haemorrhagic fever (CCHF), which can cleave polyubiquitin chains (shown in **h** as control for activity), and a mutant (OTU (2A)) with reduced activity, were used to cleave ubiquitinated E of ZIKV. After incubation of ZIKV with purified recombinant OTU, the ability of the deubiquitinated virus to attach to cells and to replicate was tested by incubation with JEG-3 cells at 4 °C for 30 min and viral RNA quantified by qPCR (**i**), and replication by plaque assay (**j**). Representative of two independent experiments, $n = 3$ technical replicates, mean ± s.e.m., unpaired two-sided t-test, $^{**}P < 0.01$.

natureresearch

Corresponding author(s): Ricardo Rajsbaum

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Zeiss LSM 880 and BD LSRII Fortessa Analyzer (BD Biosciences; San Jose, CA, USA)
Data analysis	ImageJ v1.51n, BD FACSDiva 8.0.1 software, FlowJo v9.3.2, and GraphPad Prism v7.03. Thermo Proteome Discoverer version 1.2.0.208, MaxQuant Version 1.6.8.0, and EMAN v2.02

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data is available in the main text or the supplementary materials. Mutant viruses, knockout cell lines and knockout mice may be available upon request after publication and respective material transfer agreements are completed.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	the sample size was sufficient based on power calculations and differences observed between groups giving statistically significant analysis, and based on reproducibility between experiments. Low observed variability between samples and differences observed using internal controls. Neutralization assays did not have sample size estimate and sample size relied on statistical significant data.
Data exclusions	No data were excluded
Replication	Most experiments were performed at least 2 times giving consistent and reproducible results. Experiments were performed using multiple different approaches that helped reach the same conclusions.
Randomization	Mice of matched age, sex and samples were randomly allocated to the experiments.
Blinding	Collection of samples was not blinded. However, samples were analyzed blinded and group allocation was performed afterwards.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Unique biological materials
☐	☒	Antibodies
☐	☒	Eukaryotic cell lines
☒	☐	Palaeontology
☐	☒	Animals and other organisms
☒	☐	Human research participants

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☐	☒	Flow cytometry
☒	☐	MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	The mutant Zika viruses, TRIM7 knockout cells and mice may be available upon request after publication and respective material transfer agreements are completed and costs of transfer are covered.

Antibodies

Antibodies used	Rabbit anti-E antibody (Catalog # GTX133314, Lot: 43306, 1:1000), rabbit anti-capsid (Catalog # GTX133317, Lot: 42564, 1:1000), and rabbit PrM (Catalog # GTX133305, 1:1000) were from Genetex, rabbit anti-HA antibody (Catalog # H6908, Lot: 098M4812, 1:1000), anti-FLAG antibodies (Catalog # F1804, Lot: 078M4886, 1:1000) and rabbit anti-β-actin (Catalog # A2066, Lot: GR3244114-1) antibodies were from Sigma. Rabbit monoclonal anti-ubiquitin Lysine 48 (K48, clone Apu2) (Catalog # 05-1407, Lot 3065576 and 3091725, 1:1000) and mouse monoclonal anti-ubiquitin Lysine 63 (K63, clone Apu3) (Catalog # 05-1308, Lot: 3243379) were purchased from Millipore. Rabbit polyclonal TRIM7 was from Abgent and Sigma (AP11979a and HPA039213 respectively), anti-Flavivirus group antigen antibody from Millipore (MAB10216) and mouse IgG1, κ Isotype Control from BD Biosciences (555746), and mouse total Ubiquitin UbP4D1 (Catalog # BML-PW0930-0100, Lot: 06031941, 1:1000) from Enzo, rabbit UbcH5a/UBE2D1 Antibody (Catalog # NBP1-32734Lot: 40023, 1:1000) from Novus. Secondary antibodies used ECL anti-Rabbit IgG, Horseradish peroxidase linked species-specific whole antibody (from donkey) (Catalog # NA934, Lot: 16953209, 1:10000) and ECL anti-mouse IgG, Horseradishperoxidase linked species-specific whole antibody (from sheep) (Catalog # NA931, Lot:16953154, 1:10000) from GE Healthcare. The following fluorescently labeled secondary antibodies were used for imaging: WGA-FITC (Catalog # GTX01502, Lot 821801629, 1:100) from Genetex, Alexa Fluor 555 donkey anti-rabbit (Catalog # A31572, Lot: 714269, 1: 200), Alexa Fluor 633 goat anti-mouse (Catalog # A21050, Lot: 1661234, 1:200), and DAPI (Catalog # D1306,

	1:200) were purchased from ThermoFisher.
Validation	some general antibodies, including Flag, HA, Ubiquitin, have been validated in previous publications: Xie, X.; Zou, J.; Zhang, X.; Zhou, Y.; Routh, A.L.; Kang, C.; Popov, V.L.; Chen, X.; Wang, Q.Y.; Dong, H.; et al. Dengue NS2A Protein Orchestrates Virus Assembly. Cell Host Microbe 2019, 26, 606–622. Bharaj P., Atkins C., Luthra P., Giraldo M.I., Dawes B.E., Miorin L., Johnson J.R., Krogan N.J., Basler C.F., Freiberg A.N., Rajsbaum R. The host E3-ubiquitin ligase TRIM6 ubiquitinates the Ebola virus VP35 protein and promotes virus replication. J. Virol. 2017;(18). M. Giraldo, O. Vargas-Cuartas, J.C. Gallego-Gomez, P.Y. Shi, L. Padilla-Sanabria, J.C. Castaño-Osorio, et al. K48-linked polyubiquitination of dengue virus NS1 protein inhibits its interaction with the viral partner NS4B Virus Res., 246 (2018), pp. 1-11. Validation of all antibodies against viral proteins were validated by the company (Genetex), using lysates from infected and non-infected cells. TRIM7 antibody was validated in our study using lysastes from WT and TRIM7 knockout cells, and by SIgma and Abgen.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T (CRL-11268), Huh-7, U-118 MG (HTB-15), Vero (CCL-81), JEG-3 (HTB-36),A549 (CCL-185) HTR-8 (CRL-3271), and 15P-1, cell lines were obtained from ATCC.
Authentication	Cells were not further authenticated. TRIM7 CRISPR KO cell lines were corroborated using anti-TRIM7 antibody by immunoblot
Mycoplasma contamination	Cell lines were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	U-118 MG is a brain glyoblastoma cell line, listed in ICLAC. The cell line was obtained by ATCC and verified by them. We used this cell line to test ZIKV replication in a brain cell line. We used this cell line only as an example of ZIKV replication in a brain cell line. These results were validated in vivo in brain tissue, upon infection of mice.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mouse experiments were performed in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocols were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Texas Medical Branch or at Rocky Mountain laboratories of the NIH/NIAID. Mice were maintained in a specific-pathogen-free environment according to the University of Texas Medical Branch guidelines. Three to 4-week old A129 (Ifnar1-/-), males and females were used. AB6 Ifnar1−/− (C57BL/6 mice lacking type I IFN receptor) and Tim1−/− Ifnar1−/− mice were infected at 3 weeks of age, males and females were used. Trim7 knockout mice (Trim7-/-), C57BL/6 background, 4-5 weeks of age, males and females were used.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve field-collected samples.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were carefully washed twice with PBS and trypsinized for 2min at 37°C. The cell suspension was washed twice with cold 10% FBS in PBS to remove trypsin and kept on ice until resuspended in 8% PFA at 4°C for 20 min. Samples were then washed twice with cold PBS and centrifuged (1200 rpm 4°C for 5 min).
Instrument	BD LSRII Fortessa Analyzer (BD Biosciences; San Jose, CA, USA) equipped with a 488-nm laser (495 nm/519 nm)
Software	Controlled by BD FACSDiva 8.0.1 software and used FlowJo software v9.3.2(Tree Star) for analysis.

Cell population abundance	The percentages of cell populations for representative samples of each experimental group and controls are shown in Extended Data Figs 10-11.
Gating strategy	Cells were gated by physical parameters using FSC-A and SSC-A then target cells were analyzed with FITC (FL-1) channel. A second gate was used to discriminate the autofluorescence background from positive DIOC18 signal using mock control samples and cells infected with DIOC18-stained ZIKV. the gating strategy is shown in supplementary information Figure 3

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Violet-light suppression of thermogenesis by opsin 5 hypothalamic neurons

https://doi.org/10.1038/s41586-020-2683-0

Received: 19 February 2020

Accepted: 4 August 2020

Kevin X. Zhang[1,2,3,4], **Shane D'Souza**[1,2,3], **Brian A. Upton**[1,2,3,4], **Stace Kernodle**[5], **Shruti Vemaraju**[1,2], **Gowri Nayak**[1,2], **Kevin D. Gaitonde**[1,2,3,4], **Amanda L. Holt**[6], **Courtney D. Linne**[1,2,3,4], **April N. Smith**[1,2], **Nathan T. Petts**[7], **Matthew Batie**[7], **Rajib Mukherjee**[8], **Durgesh Tiwari**[9], **Ethan D. Buhr**[10], **Russell N. Van Gelder**[10,11,12], **Christina Gross**[9,13], **Alison Sweeney**[6], **Joan Sanchez-Gurmaches**[8,13,14], **Randy J. Seeley**[5,15] & **Richard A. Lang**[1,2,14,16] ✉

The opsin family of G-protein-coupled receptors are used as light detectors in animals. Opsin 5 (also known as neuropsin or OPN5) is a highly conserved opsin that is sensitive to visible violet light[1,2]. In mice, OPN5 is a known photoreceptor in the retina[3] and skin[4] but is also expressed in the hypothalamic preoptic area (POA)[5]. Here we describe a light-sensing pathway in which POA neurons that express *Opn5* regulate thermogenesis in brown adipose tissue (BAT). We show that *Opn5* is expressed in glutamatergic warm-sensing POA neurons that receive synaptic input from several thermoregulatory nuclei. We further show that *Opn5* POA neurons project to BAT and decrease its activity under chemogenetic stimulation. *Opn5*-null mice show overactive BAT, increased body temperature, and exaggerated thermogenesis when cold-challenged. Moreover, violet photostimulation during cold exposure acutely suppresses BAT temperature in wild-type mice but not in *Opn5*-null mice. Direct measurements of intracellular cAMP ex vivo show that *Opn5* POA neurons increase cAMP when stimulated with violet light. This analysis thus identifies a violet light-sensitive deep brain photoreceptor that normally suppresses BAT thermogenesis.

The availability of photons emanating from our sun has been exploited for adaptive advantage by almost all living systems. For example, the visual sense of animals relies on detection of radiant photons for object identification. Plants and animals also anticipate the daily light–dark cycle using non-visual pathways to entrain circadian clocks. In animals, both visual and non-visual pathways use the eyes for photic input, but extraocular light detection has been well-described in non-mammalian species. For example, in the fruitfly and in zebrafish, light can entrain the circadian clock in organs directly, without the need for input from the eyes[6]. Although it had been thought that mammals do not use extraocular light detection, this view has recently changed[4,7–9].

In animals, most light-response pathways use a member of the opsin family of G-protein-coupled receptors as a light detector. Of the non-visual opsins, melanopsin (OPN4), a blue-light-sensitive (480 nm λ_{max}) opsin, has been most extensively studied in mice: ocular melanopsin has a role in circadian entrainment[10], the pupillary light reflex[11], eye development[12] as well as mood and learning[13]. Recently, studies of the visual violet-light-sensitive (380 nm λ_{max})[1,2] neuropsin (OPN5) and blue-light-sensitive encephalopsin (OPN3) have provided evidence for their involvement in extraocular light response pathways. In birds, expression of *Opn5* in the brain is implicated in the regulation of seasonal breeding behaviour[14] and in mice is necessary and sufficient for direct photoentrainment of retinal, corneal and skin circadian clocks[3,4]. OPN3 was recently shown to be expressed in adipocytes, where it promotes lipolysis in a blue-light-dependent manner[7,15].

In the mouse and primate hypothalamus, *Opn5* is expressed in the POA[5] and this raised the possibility that, as in birds, OPN5 might function as a deep brain light sensor. The POA is a thermoregulatory region that in mouse modulates the heat-generating capacity of BAT via sympathetic nervous system activity[16]. Homeotherms rely on this system to defend core body temperature against ever-changing environments. This hypothalamic–BAT neuraxis has been extensively described[17]. Here, we provide evidence that the thermoregulatory apparatus of mice is responsive to violet light in an OPN5-dependent manner. We further show that the crucial light-sensitive cells are neurons that reside in the preoptic area of the hypothalamus.

---

[1]The Visual Systems Group, Abrahamson Pediatric Eye Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [2]Center for Chronobiology, Division of Pediatric Ophthalmology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [3]Molecular and Developmental Biology Graduate Program, University of Cincinnati, College of Medicine, Cincinnati, OH, USA. [4]Medical Scientist Training Program, University of Cincinnati, College of Medicine, Cincinnati, OH, USA. [5]Department of Surgery, University of Michigan, School of Public Health, Ann Arbor, MI, USA. [6]Department of Physics, Yale University, New Haven, CT, USA. [7]Division of Clinical Engineering, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [8]Division of Endocrinology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [9]Division of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [10]Department of Ophthalmology, University of Washington Medical School, Seattle, WA, USA. [11]Department of Biological Structure, University of Washington Medical School, Seattle, WA, USA. [12]Department of Pathology, University of Washington Medical School, Seattle, WA, USA. [13]Department of Pediatrics, University of Cincinnati, College of Medicine, Cincinnati, OH, USA. [14]Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [15]Department of Internal Medicine, University of Michigan, School of Public Health, Ann Arbor, MI, USA. [16]Department of Ophthalmology, University of Cincinnati, College of Medicine, Cincinnati, OH, USA. ✉e-mail: Richard.Lang@cchmc.org

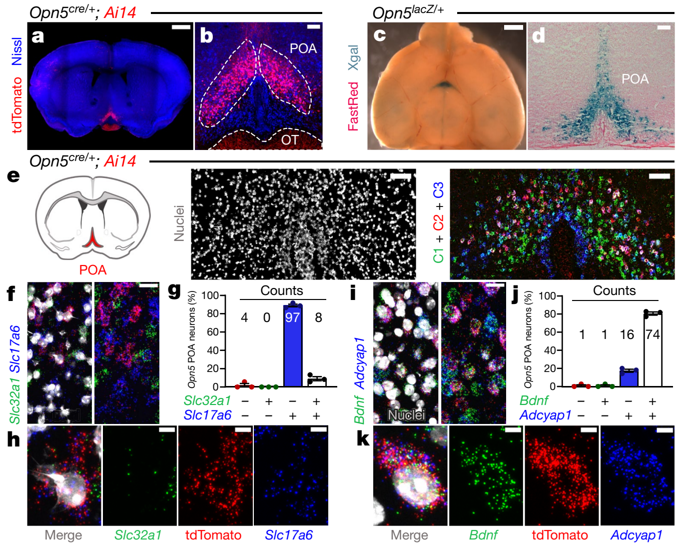


Fig. 1 | *Opn5* is expressed in excitatory, warm-sensitive hypothalamic POA neurons. a, **b**, Coronal brain section (P21 *Opn5*$^{cre/+}$*;Ai14*) showing *Opn5* (tdTomato, red) restricted to the POA. Nissl labelling is blue. Red labelling in optic tracts (OT) are axons from *Opn5* retinal ganglion cells. **c**, **d**, Xgal labelling (P10 *Opn5*$^{lacZ/+}$) in whole brain, ventral view (**c**) and coronal section through POA (**d**). **e**, Schematic of the M-FISH (Methods) region and low-magnification images of nuclear (DAPI, greyscale) and three-colour probe labelling. **f**–**h**, M-FISH for *tdTomato* (*Opn5*^{cre}*;Ai14*, red), *Slc32a1* (green), and *Slc17a6* (blue), low- (**f**) and high- (**h**) magnification images with quantification (**g**) ($n = 3$, 109 cells). **i**–**k**, M-FISH for *tdTomato* (*Opn5*^{cre}*;Ai14*, red), *Bdnf* (green), and *Adcyap1* (blue), low- (**i**) and high- (**k**) magnification images with quantification (**j**) ($n = 3$; 92 cells). **k**, Representative tdTomato$^+$ cell from **i**. Scale bars, 5 µm (**h**, **k**), 20 µm (**f**, **i**), 75 µm (**e**), 100 µm (**b**), 150 µm (**d**), 1 mm (**a**, **c**). Data in **g** and **j** are mean ± s.e.m.

Opn5 in POA thermoregulatory neurons

Using an *Opn5*^{cre} knock-in allele to activate the tdTomato reporter *Ai14* (*Opn5*$^{cre/+}$*;Ai14* mice), we identified *Opn5* expression in the POA of the hypothalamus in postnatal day (P) 21 mice (Fig. 1a, b). We confirmed that this region was actively transcribing *Opn5* using Xgal labelling in brain tissue from P10 *Opn5*$^{lacZ/+}$ mice (Fig. 1c, d). Ai14$^+$ neurons were also found in the raphe pallidus (Extended Data Fig. 1a–c) but were Xgal-negative in cryosections from P12 *Opn5*$^{lacZ/+}$ mice (Extended Data Fig. 1d), which suggests *Opn5*$^{cre/+}$*;Ai14* lineage marking from an earlier developmental stage. A comprehensive lineage survey outside the central nervous system revealed no expression of *Opn5* in brown and white adipose tissue, thyroid, liver, heart, adrenal glands and pancreas (Extended Data Fig. 1e–k).

The POA contains several discrete neuronal subtypes associated with homeostatic control. We used multiplex fluorescence in situ hybridization (M-FISH) to label distinct subpopulations in the POA of P21 *Opn5*$^{cre/+}$*;Ai14* mice (Fig. 1e): most *Opn5* POA neurons expressed *Slc17a6* (encoding VGLUT2, vesicular glutamate transporter 2) and thus were glutamatergic, whereas only a small fraction colabelled with *Slc32a1* (VGAT, vesicular GABA transporter) (Fig. 1f–h). The POA also contains temperature-sensitive neurons that co-express the neuropeptides PACAP (encoded by *Adcyap1*) and BDNF (*Bdnf*)[18] and use TRPM2 as a heat sensor[19]. Using M-FISH, we found that nearly all *Opn5* POA neurons colabelled for *Adcyap1* and *Bdnf* (Fig. 1i–k), and approximately half co-express *Trpm2* (Extended Data Fig. 2a–c). Thus, *Opn5* POA neurons are BDNF$^+$ PACAP$^+$ warm-sensitive glutamatergic neurons.

To map presynaptic inputs to *Opn5* POA neurons, we injected a tracing rabies virus into the POA of P21 *Opn5*$^{cre/+}$*;Ai6;RΦGT*[20] mice (Extended Data Fig. 2d, e). Six days after injection, we identified tdTomato-positive neurons in the paraventricular nucleus (Extended Data Fig. 2f-h), the supraoptic nucleus (Extended Data Fig. 2f, g, i), the dorsomedial hypothalamus (Extended Data Fig. 2j, k), the lateral parabrachial nucleus (Extended Data Fig. 2l, m) and the raphe pallidus (Extended Data Fig. 2n, o). These regions all have a role in thermoregulation (Extended Data Fig. 2p), with the dorsomedial hypothalamus, lateral parabrachial nucleus and raphe pallidus directly implicated in the cutaneous thermosensory circuit that controls BAT activity[17]. Together, these results indicate that *Opn5* POA neurons are an excitatory, warm-sensitive population that are synaptically connected to thermoregulatory nuclei.

Opn5 POA neurons regulate BAT activity

We next evaluated whether *Opn5* POA neurons communicated with BAT. We injected a transneuronal retrograde pseudorabies virus (PRV) expressing monomeric red fluorescent protein (mRFP1) into the BAT of P60 *Opn5*$^{cre/+}$*;Ai6* mice (Fig. 2a). Five days after injection, we identified mRFP1-positive neurons in the intermediolateral nucleus of the spinal cord, raphe pallidus, dorsomedial hypothalamus, paraventricular nucleus, the nucleus tractus solitarius and the lateral hypothalamic area (Fig. 2b–g), all regions that are implicated in BAT thermogenesis. Notably, we identified mRFP1-positive neurons in the POA that colabelled with *Ai6* (Fig. 2h, i), demonstrating that a direct polysynaptic pathway exists between *Opn5* POA neurons and the BAT.

To determine whether *Opn5* POA neurons can control BAT activity, we used chemogenetics to activate or inhibit these neurons while monitoring BAT and core temperature. Stimulatory hM3Dq or inhibitory hM4Di designer receptors exclusively activated by designer drugs (DREADDs) were targeted to *Opn5* POA neurons by injecting a Cre-dependent adeno-associated viral vector (AAV5) into *Opn5*$^{cre/+}$ mice (*Opn5*$^{+/+}$ mice were used as a control) (Fig. 2j–m). Mice were implanted with a telemetric sensor to monitor BAT and core temperature and each received an injection of the DREADD ligand clozapine *N*-oxide (CNO) for experimental and control studies (Fig. 2m). Mice were then euthanized and the BAT was collected for molecular profiling of thermogenic gene expression (Extended Data Fig. 3a, b). We found that chemogenetic activation of *Opn5* POA neurons significantly suppressed BAT and core temperature (Fig. 2n, o). Cre-negative *Opn5*$^{+/+}$ mice administered either vehicle or CNO failed to show a similar effect (Fig. 2p, q). By contrast, chemogenetic inhibition of *Opn5* POA neurons augmented BAT and core temperature (Fig. 2r, s), with this effect absent in Cre-negative controls (Fig. 2t, u). Subsequent studies performed on *Opn5*$^{cre/-}$ (loss of *Opn5* function) mice (Extended Data Fig. 3c–f) and mice under 4-°C cold exposure (Extended Data Fig. 3g–l) showed that heterozygous *Opn5* loss of function does not change baseline BAT or core temperature (Extended Data Fig. 3g–j), and that neither loss of *Opn5* nor temperature sensing alters the chemogenetic effects of *Opn5* POA neurons on BAT activity. Overall, these results demonstrate that *Opn5* POA neurons can robustly and bidirectionally regulate BAT activity.

Elevated thermogenesis in *Opn5*-null mice

To study the function of OPN5 in thermogenesis, we used a germ-line *Opn5*-null mouse (*Opn5*$^{-/-}$)[21]. Immunodetection in *Opn5*$^{-/-}$ BAT showed increased levels of uncoupling protein UCP1 (Extended Data Fig. 4a, b) and tyrosine hydroxylase (TH), a marker for innervation of the sympathetic nervous system (Extended Data Fig. 4c–e). Cold exposure revealed that that *Opn5*$^{-/-}$ mice were better at defending their body temperature and increased the expression of thermogenesis pathway genes (Extended Data Fig. 4f, g). Telemetry sensor recording further indicted that core and BAT temperature were increased in *Opn5*-null mice even at 24 °C ambient temperature, and that these differences were not due to a dysregulated circadian rhythm (Extended Data Fig. 4h, i). By infrared thermography, P8 and P90 *Opn5*$^{-/-}$ mice exposed to cold were warmer than controls (Extended Data Fig. 4j, k). Surface temperatures in the interscapular adipose (iAT) region of P90 *Opn5*$^{-/-}$ mice (Extended

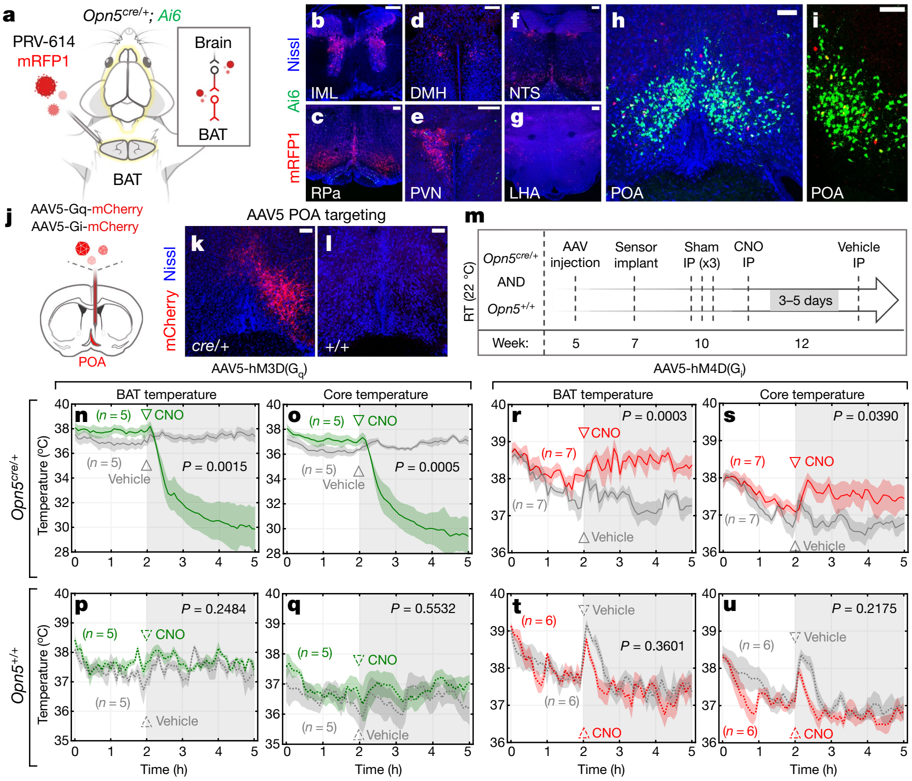


Fig. 2 | *Opn5* POA neurons regulate BAT thermogenesis. a, Pseudorabies virus (PRV-mRFP1) injection into the BAT of P60 *Opn5*$^{cre/+}$;*Ai6* mice. **b**–**i**, Representative images of PRV-infected (red) regions including the intermediolateral nucleus (IML) of the spinal cord (**b**), raphe pallidus (**c**), dorsomedial hypothalamus (**d**), paraventricular nucleus (**e**), nucleus tractus solitarius (NTS) (**f**), lateral hypothalamic area (LHA) (**g**), and *Opn5;Ai6* (green) POA neurons (**h**, **i**). **j**, Schematic of DREADD virus delivery into the POA of *Opn5*$^{cre/+}$ or *Opn5*$^{+/+}$ mice. **k**, **l**, Immunofluorescence showing AAV-infected POA neurons in *Opn5*$^{cre/+}$ mice (**k**) but not in *Opn5*$^{+/+}$ control mice (**l**). **m**, Experimental timeline. IP, intraperitoneal; RT, room temperature. **n**–**u**, Chemogenetic manipulation of *Opn5* POA neurons. CNO or vehicle (saline) injected after 2 h (open arrowhead). CNO-mediated activation of *Opn5* POA neurons with G_q DREADD decreases BAT and core temperature in *Opn5*$^{cre/+}$ mice (**n**, **o**) but not in *Opn5*$^{+/+}$ controls (**p**, **q**). CNO-mediated inhibition of *Opn5* POA neurons with G_i DREADD increases BAT and core temperature in *Opn5*$^{cre/+}$ mice (**r**, **s**) but not in controls (**t**, **u**). Scale bars, 100 μm. Data in **n**–**u** are mean ± s.e.m. All *P* values represent one-way repeated-measures analysis of variance (ANOVA).

Data Fig. 4l) were quantifiably warmer, whereas tail temperatures were indistinguishable (Extended Data Fig. 4m). In aggregate, these data suggested that mice lacking *Opn5* exhibit increased BAT thermogenesis.

The exaggerated thermogenesis of *Opn5*$^{-/-}$ mice does not lead to changes in body weight and composition or locomotor activity but does result in increased energy expenditure (Extended Data Fig. 5a–f). Lack of differences in body composition may be explained by the increased food and water consumption of *Opn5*$^{-/-}$ mice (Extended Data Fig. 5g, h). Serum lipids are lower in the *Opn5*-null, but serum levels of thyroxine (T4) and thyrotropin-releasing hormone (TRH) are unchanged (Extended Data Fig. 6a–f), suggesting that facultative and not obligatory thermogenesis is primarily affected. Major white adipose depots are smaller in *Opn5*$^{-/-}$ mice, and show decreased adipocyte size and increased levels of UCP1 (Extended Data Fig. 6g–i). Systolic and diastolic blood pressure, mean arterial pressure, and pulse rate are not different between *Opn5*$^{-/-}$ and *Opn5*$^{+/+}$ mice (Extended Data Fig. 6j–l). However, *Opn5*$^{-/-}$ mice show an augmented response to the β_3-adrenergic agonist CL-316,243 (Extended Data Fig. 6m). These results indicate that the exaggerated BAT thermogenesis of *Opn5*$^{-/-}$ mice cannot be attributed to differences in thyroid hormone or cardiovascular activity, but instead is explained by adaptive changes in adrenergic BAT sensitivity and lipid mobilization. A POA-specific *Opn5* deletion was generated using *Lepr*^{cre} mice (Extended Data Figs. 4n, o, 7a–g). We repeated the previous analyses on control (*Opn5*$^{fl/fl}$) and conditional mutant (*Lepr*$^{cre/+}$;*Opn5*$^{fl/fl}$) mice and found that mutant mice largely phenocopied the global *Opn5* loss-of-function model (Extended Data Fig. 4p–v). These data provide strong support for a BAT thermogenic-suppressive role of preoptic OPN5.

Violet light suppresses BAT activity

The observation that *Opn5*$^{-/-}$ mice show an exaggerated thermogenic response suggested that OPN5 normally inhibits thermogenesis. To assess whether this suppressive role depends on the light-sensing function of OPN5, we monitored BAT and core temperature in cold-exposed P90–P120 *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice while providing acute 380-nm violet light stimulation. In *Opn5*$^{+/+}$ mice, violet photostimulation decreased BAT and core temperatures, whereas *Opn5*$^{-/-}$ mice failed to respond (Fig. 3a, b). When violet light was not supplemented, there was no longer any divergence in BAT and core temperature between *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice (Fig. 3c, d). To assess the possibility that the addition of violet light might invoke a differential behavioural response in *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice, locomotor activity was recorded and revealed no differences in average speed or distance travelled (Extended Data Fig. 8a–i).

Opn5 is expressed in retinal ganglion cells and can photoentrain a retinal circadian clock[3]. We used two approaches to assess the possibility that retinal OPN5 might contribute to changes in BAT thermogenesis. First, we conditionally deleted *Opn5*^{fl} from retinal progenitors using *Rx*^{cre}, and found no differences in core temperature between cold-exposed wild-type (*Opn5*$^{fl/fl}$) and retinal *Opn5* conditional (*Rx*^{cre};*Opn5*$^{fl/fl}$) mice (Fig. 3e). Second, we enucleated P90–P120 *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice and subjected them to the same cold-exposure photostimulation assay as sighted mice. Enucleated *Opn5*$^{+/+}$ mice decreased their core temperature in response to violet light, whereas enucleated *Opn5*$^{-/-}$ mice showed no such response (Fig. 3f, g). Molecular profiling

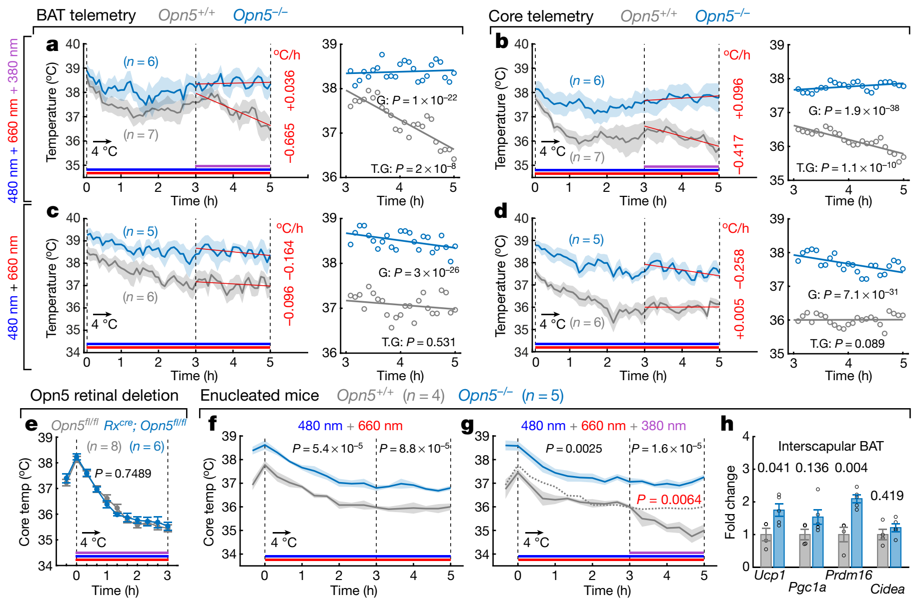


Fig. 3 | Violet light acutely suppresses BAT thermogenesis. a–**d**, BAT and core telemetry recordings during 5 h exposure to 4 °C with modulation of lighting wavelength. All mice received 480-nm and 660-nm light exposure (Methods). At the 3-h mark (dotted line), *Opn5*$^{+/+}$ or *Opn5*$^{-/-}$ mice were either supplemented with 380-nm light (**a**, **b**) or remained in 480 nm + 660 nm (**c**, **d**). BAT and core temperature trajectories during light modulation (hours 3–5) were calculated via linear regression and the rate of temperature change reported as °C per h. **e**, Core temperature assessment (rectal) of *Opn5*$^{fl/fl}$ and *Rx*^{cre};*Opn5*$^{fl/fl}$ mice during 3 h cold challenge in 380 nm + 480 nm + 660 nm lighting. **f**, **g**, Core temperature assessment in enucleated *Opn5*$^{+/+}$ ($n = 4$) and *Opn5*$^{-/-}$ ($n = 5$) mice under 480 nm + 660 nm illumination (**f**) or supplemented with 380-nm violet light (**g**) at hour 3 (dotted line). Dotted trace in **g** represents wild-type average trace from **f**. **h**, iBAT qPCR of thermogenesis genes (*Ucp1*, *Pgc1a*, *Prdm16* and *Cidea*) after 5 h cold exposure in mice from **g**. Data are mean ± s.e.m. *P* values are from one-way ANCOVA with time as covariate (for genotype (G), or time and genotype (T.G.)) (**a**–**d**), one-way repeated measures ANOVA (**e**–**g**), or ANOVA with Tukey post-hoc analysis (**h**).

of dissected BAT from enucleated *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice showed differences in thermogenic gene induction (Fig. 3h) that resembled changes observed in sighted mice. These data show that the inhibitory role of OPN5 on BAT thermogenesis does not require retinal OPN5.

Violet light absence enhances BAT activity

As an extension of our acute response analysis, we asked whether chronic elimination of violet photons would mimic *Opn5* loss-of-function in wild-type mice. Male and female wild-type mice on a C57BL6/J background were raised under 'full spectrum' (380 nm + 480 nm + 660 nm) or 'minus violet' (480 nm + 660 nm) lighting from embryonic day (E) 16.5 to P70 under a standard 12 h/12 h light–dark cycle (Extended Data Fig. 9a, b). Analysis at P70 (Extended Data Fig. 9c–m) revealed that the minus-violet mice showed a milder version of the exaggerated thermogenesis phenotype characteristic of the *Opn5*-null mice.

Because our aggregated data suggested that *Opn5* POA neurons might be directly light responsive, we assessed whether the POA received sufficient photon flux for opsin activation. Using a custom-designed optic fibre probe, we performed intra-tissue radiometry at various depths in the brain of anaesthetized mice (Extended Data Fig. 10a–c). At the λ_{max} of the OPN5 action spectrum, we measured an approximately 2.5 log-transformed fold intensity attenuation relative to the cranial surface at the depth of the POA (Extended Data Fig. 10d, e). When extrapolating for normal sunlight intensities, a maximum violet flux of 9.0×10^{12} photons cm^{-2} s^{-1} can reach the POA. This is above the activation threshold for other mammalian nonvisual opsins[22].

Opn5 POA neurons respond to violet light

Our findings raised the crucial question of how *Opn5* POA neurons signal in response to violet light. To gain insight into these mechanisms, we monitored real-time intracellular cyclic AMP (cAMP) using a genetically encoding TEpacVV cAMP sensor activated transcriptionally with *Opn5*^{cre}. TEpacVV reports cAMP binding by changes in fluorescence resonance energy transfer (FRET) between an mTurquoise donor (CFP) and a Venus acceptor (^{cp173}Venus-Venus, YFP)[23] that can be imaged using two-photon microscopy (Fig. 4a, b). Neurons that experience an increase in intracellular cAMP, such as the response to forskolin and 3-isobutyl-1-methylxanthine (IBMX), will have an increase in the ratio of CFP to YFP (Δ*F*), whereas depleting cAMP by permeabilizing the cell with digitonin will decrease Δ*F* (Fig. 4c–e). We designed a 1-h experimental protocol in which 15 min of FRET measurements in darkness were followed by 30 min of 50% duty cycle violet photostimulation with measurements taken in between, ending with 15 min of dark measurements following the application of forskolin plus IBMX (Fig. 4f). POA slices from P21 *Opn5*$^{cre/+}$ mice showed a marked increase in relative Δ*F* in response to violet photostimulation, whereas slices from P21 *Opn5*$^{cre/-}$ mice featured little to no increases in Δ*F* and were indistinguishable from dark conditions (Fig. 4g–k). These data indicate that *Opn5* POA neurons are directly sensitive to violet photostimulation ex vivo and in response, increase intracellular cAMP.

Discussion

We present evidence in mice of a violet light-sensitive thermoefferent pathway from POA to BAT that uses OPN5 as a light sensor. Acting as a deep brain photoreceptor with a peak sensitivity of 380 nm, OPN5 inhibits BAT thermogenesis through a direct light response that raises intracellular levels of cAMP.

Deep brain photoreceptors have been extensively documented in teleost[24] and avian species[14], in which nonvisual opsins regulate a host of behavioural and reproductive responses. By contrast, evidence of extraocular light sensing in mammals has only recently gained acceptance, with the precise signalling mechanisms not yet fully determined. It was previously demonstrated that adipocyte

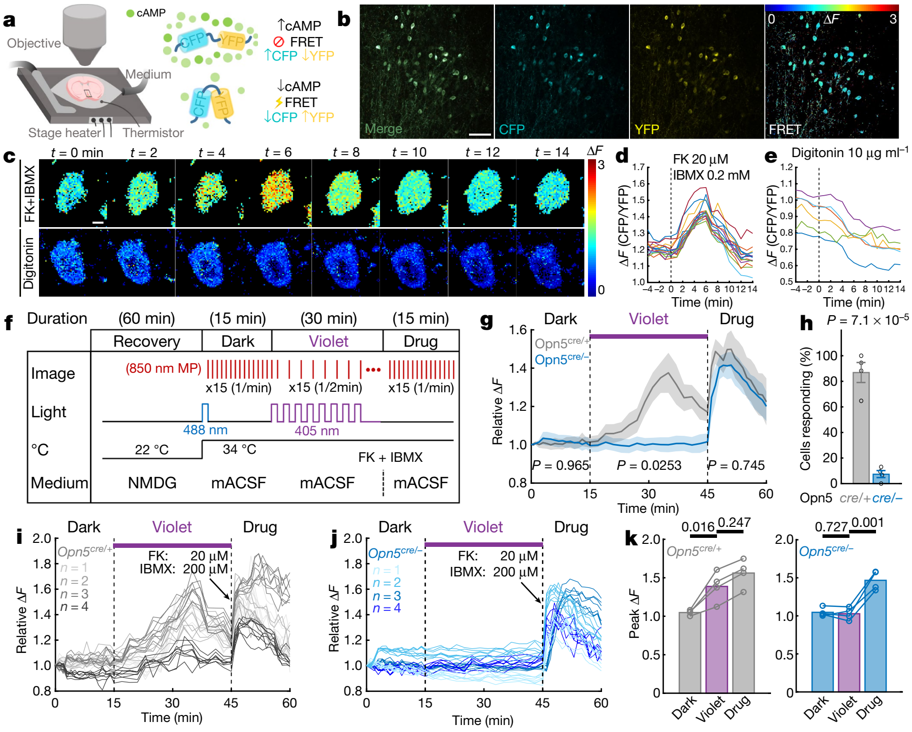


Fig. 4 | *Opn5* POA neurons respond to violet light ex vivo. a, Schematic depicting two-photon assessment of cAMP biosensor FRET activity in POA slices from *Opn5*^{cre};*CAMPER* mice. **b**, CFP, YFP and FRET images (expressed as Δ*F* = CFP/YFP ratio). **c**, Time course Δ*F* images after response to forskolin (FK, 20 μM) and IBMX (200 μM) (top) or digitonin (10 μg ml^{-1}) (bottom). **d**, **e**, Individual traces from slices treated with FK + IBMX (**d**, *n* = 15 cells) or digitonin (**e**, *n* = 6 cells). **f**, Experimental timeline for testing violet responses of *Opn5* neurons in POA slices as described in Methods. **g**, Relative Δ*F* plots for *Opn5*$^{cre/+}$ (grey trace, *n* = 4) and *Opn5*$^{cre/-}$ (blue trace, *n* = 4) mice. **h**, Percentage of cells responding to violet light (average relative Δ*F* > 1.1 between *t* = 15 and *t* = 45) for both groups. **i**, **j**, Individual traces from each biological replicate (*n* = 6–8 cells per mouse, 4 mice per genotype) from experiments in **g**. **k**, Peak Δ*F* from dark, violet stimulation, and drug phases between *Opn5*$^{cre/+}$ and *Opn5*$^{cre/-}$ mice. Data in **g**, **h** and **k** are presented as mean ± s.e.m. *P* values are from one-way repeated measures ANOVA (**g**), or two-tailed Student's *t*-test (**h**, **k**). Scale bars, 10 μm (**c**), 100 μm (**b**).

OPN3 (a blue-light sensitive opsin) increases lipolysis by promoting cAMP-dependent phosphorylation of hormone-sensitive lipase and thus enhances adaptive thermogenesis in mice[7]. In this report, we use neuroanatomical and loss-of-function studies to establish preoptic area OPN5 in an inhibitory role for BAT thermogenesis. Chemogenetic stimulation of *Opn5* POA neurons immediately decreases BAT temperature, whereas mice that lack *Opn5* show marked increases in adaptive thermogenesis and BAT activity. These opposing activities of OPN3 and OPN5 on thermogenesis raise the interesting hypothesis that nonvisual photoreceptive pathways decode light information to help calibrate time-of-day appropriate BAT activity. The precise mechanisms that integrate OPN3 and OPN5 activities in thermogenesis pathways require further study.

Recent functional studies on POA neurons that express BDNF, PACAP[18], leptin receptor[25], TRPM2[19] and prostaglandin EP$_3$ receptor[26] have uncovered evidence that glutamatergic, and not GABA-producing, neuronal populations directly regulate body temperature. These data challenge previous models that suggest that BAT-projecting thermoregulatory POA neurons are GABAergic[17]. Our analysis identified *Opn5* POA neurons to be glutamatergic, double-positive for BDNF and PACAP, and, consistent with the outcome of previous studies, elicited robust decreases in BAT and core temperature under chemogenetic stimulation. Several lines of evidence now suggest that an excitatory subpopulation of warm-sensitive POA neurons can integrate signalling from leptin, prostaglandin E$_2$ and violet light. The most compelling evidence of this comes from a tandem single-cell RNA-sequencing multiplexed error-robust FISH (scRNA-seq–MERFISH) cell atlas of the POA, in which excitatory subcluster e13 is enriched for co-expression of *Adcyap1*, *Bdnf*, *Slc17a6*, *Ptger3* (prostaglandin E$_2$ receptor), *Lepr* and *Opn5*[27]. It will be crucial to investigate whether this population represents a bona fide nexus for signal integration for all these pathways.

In summary, our findings have revealed an unexpected light-responsive POA–BAT neuraxis in mice that requires OPN5 as a deep brain light sensor. It will be interesting for future work to examine the possibility that normal thermogenesis in humans requires light input via extraocular pathways. This possibility is supported by the conservation of *OPN3* expression in human adipocytes, *OPN5* expression in the POA of primates[5], and many metabolic diseases that show a risk that is dependent on the season of birth[28,29], suggesting involvement of light-response pathways. We speculate that insufficient stimulation of OPN3 and OPN5 in these tissues may contribute to the growing epidemic of metabolic disease in the developed world, where artificial lighting has become the norm.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2683-0.

1. Tarttelin, E. E., Bellingham, J., Hankins, M. W., Foster, R. G. & Lucas, R. J. Neuropsin (Opn5): a novel opsin identified in mammalian neural tissue. *FEBS Lett.* **554**, 410–416 (2003).
2. Kojima, D. et al. UV-sensitive photoreceptor protein OPN5 in humans and mice. *PLoS ONE* **6**, e26388 (2011).
3. Buhr, E. D. et al. Neuropsin (OPN5)-mediated photoentrainment of local circadian oscillators in mammalian retina and cornea. *Proc. Natl Acad. Sci. USA* **112**, 13093–13098 (2015).

4. Buhr, E. D., Vemaraju, S., Diaz, N., Lang, R. A. & Van Gelder, R. N. Neuropsin (OPN5) mediates local light-dependent induction of circadian clock genes and circadian photoentrainment in exposed murine skin. *Curr. Biol.* **29**, 3478–3487.e4 (2019).
5. Yamashita, T. et al. Evolution of mammalian Opn5 as a specialized UV-absorbing pigment by a single amino acid mutation. *J. Biol. Chem.* **289**, 3991–4000 (2014).
6. Whitmore, D., Foulkes, N. S. & Sassone-Corsi, P. Light acts directly on organs and cells in culture to set the vertebrate circadian clock. *Nature* **404**, 87–91 (2000).
7. Nayak, G. et al. Adaptive thermogenesis in mice is enhanced by opsin 3-dependent adipocyte light sensing. *Cell Rep.* **30**, 672–686.e8 (2020).
8. Sikka, G. et al. Melanopsin mediates light-dependent relaxation in blood vessels. *Proc. Natl Acad. Sci. USA* **111**, 17977–17982 (2014).
9. Yim, P. D. et al. Activation of an endogenous opsin 3 light receptor mediates photo-relaxation of pre-contracting late gestation human uterine smooth muscle ex vivo. *Reprod. Sci.* **27**, 1791–1801 (2020).
10. Panda, S. et al. Melanopsin (Opn4) requirement for normal light-induced circadian phase shifting. *Science* **298**, 2213–2216 (2002).
11. Lucas, R. J. et al. Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science* **299**, 245–247 (2003).
12. Rao, S. et al. A direct and melanopsin-dependent fetal light response regulates mouse eye development. *Nature* **494**, 243–246 (2013).
13. Fernandez, D. C. et al. Light affects mood and learning through distinct retina-brain pathways. *Cell* **175**, 71–84.e18 (2018).
14. Nakane, Y. et al. A mammalian neural tissue opsin (Opsin 5) is a deep brain photoreceptor in birds. *Proc. Natl Acad. Sci. USA* **107**, 15264–15268 (2010).
15. Sato, M. et al. Cell-autonomous light sensitivity via Opsin3 regulates fuel utilization in brown adipocytes. *PLoS Biol.* **18**, e3000630 (2020).
16. Tan, C. L. & Knight, Z. A. Regulation of body temperature by the nervous system. *Neuron* **98**, 31–48 (2018).
17. Morrison, S. F., Madden, C. J. & Tupone, D. Central neural regulation of brown adipose tissue thermogenesis and energy expenditure. *Cell Metab.* **19**, 741–756 (2014).
18. Tan, C. L. et al. Warm-sensitive neurons that control body temperature. *Cell* **167**, 47–59. e15 (2016).
19. Song, K. et al. The TRPM2 channel is a hypothalamic heat sensor that limits fever and can drive hypothermia. *Science* **353**, 1393–1398 (2016).
20. Takatoh, J. et al. New modules are added to vibrissal premotor circuitry with the emergence of exploratory whisking. *Neuron* **77**, 346–360 (2013).
21. Nguyen, M. T. T. et al. An opsin 5-dopamine pathway mediates light-dependent vascular development in the eye. *Nat. Cell Biol.* **21**, 420–429 (2019).
22. Wong, K. Y. A retinal ganglion cell that can signal irradiance continuously for 10 hours. *J. Neurosci.* **32**, 11478–11485 (2012).
23. Muntean, B. S. et al. Interrogating the spatiotemporal landscape of neuromodulatory GPCR signaling by real-time imaging of camp in intact neurons and circuits. *Cell Rep.* **22**, 255–268 (2018).
24. Fernandes, A. M. et al. Deep brain photoreceptors control light-seeking behavior in zebrafish larvae. *Curr. Biol.* **22**, 2042–2047 (2012).
25. Yu, S. et al. Glutamatergic preoptic area neurons that express leptin receptors drive temperature-dependent body weight homeostasis. *J. Neurosci.* **36**, 5034–5046 (2016).
26. Machado, N. L. S., Bandaru, S. S., Abbott, S. B. G. & Saper, C. B. EP3R-Expressing glutamatergic preoptic neurons mediate inflammatory fever. *J. Neurosci.* **40**, 2573–2588 (2020).
27. Moffitt, J. R. et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
28. Boland, M. R., Shahn, Z., Madigan, D., Hripcsak, G. & Tatonetti, N. P. Birth month affects lifetime disease risk: a phenome-wide method. *J. Am. Med. Inform. Assoc.* **22**, 1042–1053 (2015).
29. Kahn, H. S. et al. Association of type 1 diabetes with month of birth among U.S. youth: The SEARCH for Diabetes in Youth Study. *Diabetes Care* **32**, 2010–2015 (2009).

Methods

Where appropriate, statistical methods were used to predetermine sample size. With the exception of imaging analysis, investigators were not blinded to allocation during experiments and outcome assessment. The experiments were not randomized.

Mice

Mice were housed in a pathogen-free vivarium maintained at an ambient temperature of 22 °C and a relative humidity of 30–70%. All pharmacological and surgical procedures were conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee at Cincinnati Children's Hospital Medical Center (protocol number 2018-0046). This study is compliant with all relevant ethical regulations regarding animal research. Genetically modified mice used in this study include: *Rx-cre*, *Ai14* (Jax stock 007914), *Ai6* (Jax stock 007906), *RΦGT* (Jax stock 024708)[20], *CAMPER* (*Rapgef3* Jax Stock 032205), *Lepr-cre* (*ObRb-cre*, Jax stock 008320), and *Opn5*$^{tm1a(KOMP)Wtsi}$ mice that were generated from C57BL/6N embryonic stem cells obtained from KOMP (embryonic stem clone ID: KOMP-HTGRS6008_A_B12-Opn5-ampicillin) as previously described[21]. In brief, the embryonic stem cells contain a genetic modification in which a *lacZ-neomycin* cassette is flanked by FRT sites, between exons 3 and 4, and a *loxP* site separates *lacZ* from the neomycin coding region. *loxP* sites also flank exon 4 of *Opn5*, allowing multiple mouse lines that can serve as reporter nulls, conditional floxed and null mice. The *Opn5*^{fl} allele was created by crossing the *Opn5*$^{tm1a(KOMP)Wtsi}$ mice to *FLPeR* (Jax stock 003946) to remove the *LacZ* cassette. The *Opn5*$^{-/-}$ line was created by crossing the *Opn5*^{fl} mice to *E2a-cre* (Jax stock 003724). The *Opn5*$^{-/-}$ line was propagated under a mixed background (C57/129/CD1/FVB). Littermate control mice were used for all experiments with the exception of C57BL/6J mice, which were reared under different lighting conditions. The *Opn5*^{cre} mice were generated in-house using CRISPR–Cas9 technology as previously described[21].

Mice were placed on a normal chow diet (29% protein, 13% fat and 58% carbohydrate kcal; LAB Diet 5010) ad libitum with free access to water. Littermate controls were used for genetic crosses and both male and female mice were included in the study unless otherwise stated. Ages of mice used include P8, P10, P12, P16, P21, P27, P35, P60, P70, P90 and P120 and are indicated in the relevant experiments.

Genotyping

Primer sequences and pairs for genotyping each of the alleles in this study are listed in Supplementary Table 1.

Lighting conditions

Mice were housed in standard vivarium fluorescent lighting (photon flux 1.62×10^{15} photons cm^{-2} s^{-1}) on a 12 h/12 h light/dark cycle except where noted. For generation of 'minus violet' mice (Extended Data Fig. 8), mice were housed in lighting chambers tuned to deliver full spectrum lighting or violet restricted lighting. For full spectrum lighting (above), light-emitting diodes (LEDs) were used to yield a comparable total photon flux of 1.642×10^{15} photons cm^{-2} s^{-1}. Spectral and photon flux information for full spectrum LED lighting: near violet (λ_{max} = 395 nm, 4.904×10^{14} photons cm^{-2} s^{-1} in the 375–435 nm range), blue (λ_{max} = 470 nm, 4.035×10^{14} photons cm^{-2} s^{-1} in the 435–540 nm range), and red (λ_{max} = 660 nm, 7.411×10^{14} photons cm^{-2} s^{-1} in the 600–700 nm range). Spectral and photon flux information for minus violet LED lighting: blue (λ_{max} = 470 nm, 7.509×10^{14} photons cm^{-2} s^{-1} in the 435–540 nm range), and red (λ_{max} = 630 nm, 9.705×10^{14} photons cm^{-2} s^{-1} in the 600–700 nm range), yielding a total of 1.736×10^{15} photons cm^{-2} s^{-1}. Photon fluxes were measured at approximately 61 cm from source and through an empty standard mouse cage. For wavelength restricted experiments, C57BL/6J mice were housed in a 12 h/12 h light/dark cycle starting in late gestation (E16.5) either in full spectrum or in minus violet. These mice are referred to in the experiments as 'full spectrum' and 'minus violet' respectively.

Viral vectors

All viruses used in these studies were obtained from the Center for Neuroanatomy with Neurotropic Viruses (CNNV), through its partner institutions at Princeton University, University of Pittsburgh, and Thomas Jefferson University. For monosynaptic tracing of *Opn5* POA neurons, the CVS-N2cΔG/EnvA-tdTomato rabies virus was used, derived from the deletion mutant CVS-N2c rabies strain produced in Neuro2A neuroblastoma cells. For BAT projection mapping, PRV614-mRFP1 was used, which is an attenuated laboratory pseudorabies strain expressing red fluorescent protein mRFP1 under CMV promoter control. For chemogenetic studies, AAV5-hSyn-DIO-hM3D(G_q)-mCherry and AAV5-hSyn-DIO-hM4D(G_i)-mCherry viruses were used. The CVS-N2cΔG rabies virus was provided by M. J. Schnell. The PRV614-mRFP1 virus was provided by L. W. Enquist. The AAV5-hSyn-DIO-hM3D(G_q)-mCherry and AAV5-hSyn-DIO-hM4D(G_i)-mCherry viruses were obtained through Addgene (plasmid 44361 and 44362 respectively).

Stereotaxic surgery

Mice were anaesthetized with ventilated isoflurane (induction: 4%, maintenance: 1–2%), and affixed to a stereotaxic frame (Stoelting). To trace preoptic *Opn5* neurons, P21 *Opn5*^{cre}*;R26*$^{RΦGT/Ai6}$ mice were injected with 0.5 µl of the CVS-N2c rabies virus (titre: 1.0×10^9 plaque-forming units (PFU) ml^{-1}) into the POA (coordinates relative to bregma: +0.40 mm AP, +0.20 mm ML, -4.00 mm DV). Six days after injection, mice (P27) were euthanized and perfused with PBS and 4% paraformaldehyde. For BAT projection mapping, P60 *Opn5*^{cre}*;R26*$^{Ai6/Ai6}$ mice were dissected to expose the interscapular adipose region. Six 50-nl nanoinjections of the PRV614-mRFP1 virus (titre: 4.9×10^9 PFU ml^{-1}) were made bilaterally into the interscapular BAD. Mice were then euthanized and perfused with PBS and 4% paraformaldehyde five days after injection. For chemogenetic studies, 4-week-old male *Opn5*$^{cre/+}$, *Opn5*$^{cre/-}$ (*Opn5* reporter null), and *Opn5*$^{+/+}$ (Cre-negative control) mice were injected with 1.0 µl AAV5-hSyn-DIO-hM3D(G_q)-mCherry or AAV5-hSyn-DIO-hM4D(G_i)-mCherry virus (titre: 7×10^{12} viral genomes (vg) ml^{-1}) into the POA (coordinates relative to bregma: +0.40 mm AP, +0.20 mm ML, -4.00 mm DV). All AAV-injected mice were given a recovery period of at least 2 weeks before further experimentation.

Chemogenetic manipulation experiments

Implanted mice were transferred to the lighting chamber that was situated in either cold (4 °C) or room temperature (22 °C) conditions for chemogenetic inhibitory hM4D(G_i) experiments, or just room temperature (22 °C) for chemogenetic stimulatory hM3D(G_q) experiments. BAT and core temperature recordings were collected every 5 min for a total of 5 h, from 10:00–15:00. Lighting conditions were maintained with red (660 nm), blue (480 nm) and violet (380 nm) for the entire 5 h. At hour 2, either CNO (1.0 mg kg^{-1} G_q DREADD, 2.0 mg kg^{-1} for G_i DREADD or vehicle (saline) was administered intraperitoneally to mice. All mice received both CNO and vehicle in separate experiments, and once telemetric recordings were complete, mice were administered CNO and killed 6 h later, with relevant tissues collected and the telemetric sensor explanted.

Thermoregulation and cold exposure assays

Core body temperature assessment upon acute cold exposure was performed as previously described[7] on *Opn5*-null (*Opn5*$^{-/-}$) and littermate controls (*Opn5*$^{+/+}$). Mice with *Opn5* conditionally deleted from the retinal progenitors (*Opn5*$^{fl/fl}$ and *Rx-cre;Opn5*$^{fl/fl}$), and *Opn5* conditionally deleted from *Lepr*-expressing neurons in the POA (*Opn5*$^{fl/fl}$ and *Lepr-cre;Opn5*$^{fl/fl}$) were also subjected to this assay. Furthermore, enucleated *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice, and 'full spectrum' and 'minus violet' reared C57BL/6J mice were also cold exposed.

P60 adult male and female littermates were separated from their home cage and individually housed in a home-built lighting chamber situated in an electronically monitored 4 °C cold room for 3 or 5 h depending on the assay. While the mouse was conscious, core body temperature was measured with a RET-3 microprobe rectal thermometer (Kent Scientific Corporation) every 20 min for the duration of the assay. Food and water were available ad libitum. The thermometer probe operator was blinded to mouse genotype and previous temperature measurements throughout the experiment. At the end of the cold exposure, mice were euthanized and relevant tissues (BAT, inguinal white adipose tissue (inWAT), pgWAT) were dissected, weighed, and snap frozen for downstream molecular profiling.

For all 3-h cold exposure assays, mice were subjected to a red (660 nm), blue (480 nm), and violet (380 nm) LED combination (RBV). For 5 h cold exposure assays, mice were initially subjected to only red (660 nm) and blue (480 nm) lighting (RB) for the first three hours. After the initial 3 h, violet light (380 nm) was then supplemented during hours 4 and 5. All 3-h cold exposure assays were performed during the mice' subjective day from 11:00–14:00. The 5 h assays were all performed from 10:00–15:00.

Telemetric temperature monitoring

P60 adult male *Opn5*$^{-/-}$ mice and wild-type littermate controls (*Opn5*$^{+/+}$), and *Opn5*^{cre}*;AAV5-hM3D(Gq)* or *AAV5-hM4D(Gi)* injected mice were implanted with indwelling telemetric sensors and subjected to a 5 h cold (4 °C) or ambient (22 °C) temperature exposure assay. *Opn5*^{cre}*;AAV5-hM3D(Gq)* mice did not undergo a cold exposure assay. In brief, mice were moved to individual housing and acclimated to a soft diet (DietGel 76A, and DietGel Recovery + 1 mg/2 oz carprofen) 3 days before the implantation surgery. On the day of the surgery, mice were anaesthetized and maintained with ventilated isoflurane, and a telemetric sensor (TTA-XS, Stellar Telemetry, TSE Systems) was subcutaneously implanted in the dorsal cavity. The sensor wirelessly communicates with an external antenna, and features two external thermistor leads, one advanced underneath the iAT (BAT temperature), and one advanced through the peritoneum to rest in the visceral cavity of the mouse (core temperature). Telemetric data were acquired using BIOPAC AcqKnowledge 5.0 software. Implanted mice were returned to individual housing and monitored for at least two weeks before experiments.

Acute violet light stimulation experiments

Implanted mice were transferred to a home-built lighting chamber that was situated either in the cold (4 °C) or in room temperature (22 °C). BAT and core temperature readings were collected every 5 min for a total of 5 h, from 10:00–15:00. Lighting conditions were either maintained with red (660 nm) and blue (480 nm) for the entire 5 h, or with violet (380 nm) light supplemented for hours 4 and 5. After the experiment, mice were either returned to light-controlled housing, or euthanized and perfused with 4% paraformaldehyde, with relevant tissues collected and the telemetric sensor explanted.

Imaging intracellular cAMP dynamics

Two-photon imaging of intracellular cAMP dynamics ex vivo in acute brain slices was performed as follows.

Acute brain slice preparation

P30–P60 *Opn5*$^{cre/+}$*;CAMPER* or *Opn5*$^{cre/-}$*;CAMPER* male and female mice were dark adapted for 4 h before tissue collection. Ice-cold modified artificial cerebrospinal fluid (mACSF; 92 mM NaCl, 2.5 mM KCl, 1.25 mM NaH_2PO_4, 30 mM $NaHCO_3$, 25 mM glucose, 20 mM HEPES, 5 mM Na-ascorbate, 3 mM Na-pyruvate, 2 mM thiourea, 10 mM $MgSO_4 \cdot 7H_2O$, 0.5 mM $CaCl_2 \cdot 2H_2O$, titrated to pH 7.22 by NaOH) was equilibrated with 95% oxygen and 5% carbon dioxide. Under dim red light, mice were anaesthetized with isoflurane, thoracotamized and transcardially perfused with oxygenated ice-cold mACSF. Brains were rapidly dissected and placed in oxygenated ice-cold mACSF. Coronal 300-µm sections were cut with a vibratome (Leica VT1000 S) and placed in a foil-covered bubbled room-temperature *N*-methyl-D-glucamine recovery solution (NMDG, 92 mM *N*-methyl-D-glucamine, 2.5 mM KCl, 1.25 mM NaH_2PO_4, 30 mM $NaHCO_3$, 25 mM glucose, 20 mM HEPES, 5 mM Na-ascorbate, 3 mM Na-pyruvate, 2 mM thiourea, 10 mM $MgSO_4 \cdot 7H_2O$, 0.5 mM $CaCl_2 \cdot 2H_2O$, 92 mM *N*-methyl-D-glucamine, titrated to pH 7.25 by HCl) for 30 min. To record intracellular cAMP dynamics, slices were transferred to a recording chamber (RC-26G, Werner Instruments) and continuously perfused with 30–34 °C oxygenated mACSF at a rate of 2.1 ml min^{-1}. To isolate responses intrinsic to hypothalamic neurons, the perfused mACSF was supplemented with 1 µM tetrodotoxin citrate (HB1035, Hello Bio) during imaging.

Brain slice imaging

Two-photon imaging of FRET was performed on a Nikon A1R upright confocal microscope using the NIS Elements Confocal software package v5.20.02. Images were acquired through a 16× dipping objective (CFI75 LWD 16X W, Nikon). mTurquoise (FRET donor) was excited by tuning a TiSapphire IR laser to 850 nm for two-photon imaging, with 470–500 nm (mTurquoise; FRET donor, CFP channel) and 525–575 nm (^{cp173}Venus-Venus; FRET acceptor, YFP channel) bandpass emission filtration. To visually locate *Rapgef3*-expressing cells, the POA was briefly exposed to blue epifluorescence of 488 nm for less than a minute. For dark-treated and drug-treated cells, images were taken every minute. For 405 nm-laser illuminated cells, images were taken every other minute, with one minute of continuous 405 nm photostimulation in between. Drugs were bath-applied at the 45 min mark of the experiment. 20 µM forskolin NKH477 (344281, EMD Millipore), 200 µM IBMX (02195262-CF, MP Biomedicals), and 10 µg ml^{-1} digitonin (D141, Sigma Aldrich) were applied according to experimental time points. ΔF (change in FRET) is presented as the ratio of donor emission to acceptor emission (CFP/YFP). Images were processed and quantified using NIS Elements AR v5.20.00, ImageJ Ratio Plus plugin, and MATLAB 2018a.

Indirect calorimetry

Male and female *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice aged P90–P120 were acclimated in metabolic chambers (PhenoMaster, TSE Systems) for 3 days before the start of the study. Mice were continuously recorded for a total of 16 days with the following measurements taken every 15 min: gas exchange (O_2 and CO_2), food intake, water intake, and spontaneous locomotor activity (in the x–y plane). Ambient temperature was adjusted via climate-controlled chambers that housed the metabolic chambers. VO_2, VCO_2 and energy expenditure were calculated according to the manufacturer's guidelines (PhenoMaster Software, TSE Systems), with energy expenditure estimated via the abbreviated Weir formula. The respiratory exchange ratio (RER) was calculated by the ratio VCO_2/VO_2. Mass-dependent variables (VO_2, VCO_2, energy expenditure) were not normalized to body weight. Food and water intake were measured by top-fixed load cell sensors, from which food and water containers were suspended into the sealed cage environment. For food consumption, mice demonstrating excessive food grinding behaviour were excluded from statistical analyses. After 8 days of continuous recording, cages were replaced with clean ones and sealed, and gas exchange re-equilibration completed all within 4 h. Body mass composition (fat and lean mass) were measured using nuclear magnetic resonance and expressed as grams of fat and lean tissue, and as a percentage of total body mass.

For CL-316,243 experiments, mice aged P90–P120 were acclimated in metabolic chambers (Promethion, Sable Systems International) for 3 days before the start of the study. Male and female *Opn5*$^{+/+}$, *Opn5*$^{-/-}$, full spectrum, and minus violet mice were included for these studies. Oxygen consumption (VO_2), carbon dioxide production (CO_2), energy expenditure, RER, and locomotor activity (cm s^{-1}) were recorded every 5 min using Sable Systems International Metascreen software v.2.3.15.11. Food and water were available ad libitum. Then, 1.0 mg kg^{-1} CL316,243 or vehicle (saline) was intraperitoneally injected at hour 1 of a 6-h measurement window between 11:00 and 17:00. All mice received both CL316,243 and vehicle injections in randomized order. Data were exported using Sable Systems International ExpeData software v.1.9.27.

Infrared thermography

For whole-body infrared thermographic imaging, adult (P90) and neonatal (P8) *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice were individually housed and placed in a home-built lighting chamber situated in 4 °C for 30 min. Infrared thermographic images were taken with a FLIR T530 infrared camera (FLIR Systems) every minute for a total of 30 images per P90 adult or pair of P8 pups. To quantify interscapular region temperature, a pixel average from a region of interest drawn over the iAT was taken per image per mouse using FLIR Tools Desktop software v5.13.18031.2002. The size of the selected region of interest did not change. For the comparative infrared images, P90 adult mice were briefly anaesthetized after the 30 min cold exposure and laid side by side. To quantify surface tail temperature of adult *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice, mice were placed in a tubular mouse restraint (Kent Scientific). These restraints permitted respiration through a slotted nose cone but immobilized the mouse while exposing its tail through a rear port. Tail temperatures were quantified by describing a pixel-averaged circular region of interest of consistent size and rostrocaudal distance from the base of the tail, per minute, per mouse.

Video tracking

P60 male *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice were placed in custom built cylindrical open-top acrylic enclosures with paper bedding and enrichment situated in an electronically monitored 4 °C cold room. A recording camera (Fujifilm XT-10, with Samyang 12mm f/2.0 lens) was affixed approximately 61 cm above the cages and recorded video at 24 frames per second for a total of 140 min. Ambient 480 nm and 660 nm LEDs provided red and blue illumination, and at the 80 min mark, 380 nm violet LEDs were switched on. The video was re-encoded at 2.4 frames per second and analysed by centroid-based motion tracking in NIS Elements Ar v.5.20.00.

Noninvasive blood pressure measurements

Mice were acclimated to a tubular mouse restraint (Kent Scientific) situated on a heated stage for 2–3 days before the study. On the day of the experiment, mice were placed inside the restraint on a heated stage and connected to a tail occlusion cuff and a volume pressure recording (VPR) cuff that communicated with the CODA High Throughput Noninvasive Blood Pressure System (Kent Scientific). Thirty trials of tail occlusion and VPR recordings were automatically and sequentially gathered per mouse, and systolic/diastolic blood pressure, mean arterial pressure, and pulse rate calculated by the CODA Data Acquisition Software v.4.1.

Intra-cranial tissue radiometry

Fabrication of the Holt–Sweeney microprobe was performed as previously described[30]. The termination of one end of a 100 μm silica core fibre optic patch cable (Ocean Optics) was removed. The furcation tubing and jacketing of the fibre was stripped, and the polyimide buffer was removed 5 cm from the end of the fibre end using a butane torch. A 10 g weight was attached to the end of the fibre and then pulled upon heating with the butane torch, narrowing the diameter. The narrowed region of the fibre was then cut using carborundum paper, to yield a flat fibre end with a diameter of 30–50 μm. The sides of the narrowed fibre were painted with a film opaquing pen to prevent stray light from entering, while leaving a small transparent opening at the fibre tip. For structural support, this bare, tapered fibre was then secured in the tip of a pulled glass Pasteur pipette using a drop of cyanoacrylate glue, leaving only 6-9 mm of bare optical fibre protruding. A small light-scattering ball was added to the end of the tapered optical fibre for spectral scalar irradiance measurements. To do this, titanium dioxide was thoroughly mixed with a high-viscosity UV-curable resin, DELO-PHOTOBOND, GB368 (DELO Industrie Klebstoffe). The tip of a pulled fibre was quickly inserted and removed from a droplet of the resin and titanium dioxide mixture, resulting in a sphere with a diameter of approximately twice that of the tapered fibre. As all measurements from a given probe were normalized to the signal from the same probe in a gelatin blank, small variations in the probe diameter have no effect on our results. The sphere was cured for 12 h using a Thorlabs fibre coupled LED light source (M375F2, Thorlabs).

For intra-tissue radiometric measurements in mice, mice were anaesthetized under ventilated isoflurane and placed in a mouse stereotaxic frame (Stoelting). Hair over the scalp was shaved and the skin incised rostrocaudally to expose the skull surface. The skull was breached with a small 0.5 mm diameter micromotor drill 0.4 mm anterior and 0.2 mm lateral to bregma. Following, the Holt–Sweeney microprobe was affixed to the stereotaxic frame, positioned over AP +0.40 mm, ML +0.20 mm, and lowered to DL -4.00 mm in 0.50 mm increments. While the probe is in position, the scalp skin was repositioned to cover as much of the incision site as possible without obstructing probe descent. For broadband light illumination, a Thorlabs plasma light source (HPLS345, Thorlabs) was positioned above and in front of the mouse stereotaxic frame. The light was delivered to the mouse via a 5 mm liquid light guide connected to a 2 in. (5 cm) collimating lens secured in a vice. The distance from the collimating lens to the animal was approximately 2 ft (0.6 m).

Scalar irradiance measurements as a function of wavelength were obtained at the surface of the cortex and at probe depth increments of 0.50 mm up to 4.00 mm. Spectral irradiance data were collected using an Ocean Optics 200–850 nm spectrometer (JAZ Series, Ocean Optics) and recorded using Ocean Optics OceanView v.1.6.5 software.

Tissue processing, sectioning and immunohistochemistry

Animals were anaesthetized under isoflurane and transcardially perfused with 4% paraformaldehyde solution. For immunofluorescence, brains were dissected and post-fixed in cold 4% paraformaldehyde overnight at 4 °C. After washing in PBS, brains were cryoprotected in sucrose solution and embedded for sectioning in a cryostat (Leica CM3050 S). Thirty-micrometres μm sections were obtained and subsequently processed for immunofluorescence. For immunohistochemistry, iAT and inWAT tissues were dissected and post-fixed in 4% paraformaldehyde overnight at room temperature. After washing in PBS, tissues were processed (Leica ASP300S) and embedded (Tissue-Tek TEC 6). Embedded tissue blocks were cut using a microtome (Leica RM2255) at a thickness of 4.5 μm. Slides were incubated overnight at 4 °C in primary, rinsed, and then incubated in secondary for 1 h at room temperature. Slides were then rinsed and mounted with VectaShield HardSet antifade mounting medium with DAPI.

Antibodies used for IF include NeuroTrace 435/455 blue fluorescent Nissl stain (ThermoFisher Scientific, N21479, 1:100 dilution), anti-Isolectin IB4 antibody (ThermoFisher Scientific, I21411, 1:300 dilution), anti-Tyrosine Hydroxylase antibody (Abcam, ab113, 1:500 dilution), and anti-insulin antibody (Dako, A0564, 1:500 dilution). Antibodies used for immunohistochemistry include anti-UCP1 antibody (Abcam, ab10983, 1:500 dilution).

Xgal staining

For Xgal labelling, P10 *Opn5*^{lacZ} and P16 *Lepr-cre;Ai14; Opn5*^{lacZ} mice were anaesthetized and transcardially perfused with Xgal fixative (1% formaldehyde, 0.2% glutaraldehyde, 2 mM $MgCl_2$, 5 mM EGTA, and 0.01% Nonidet P-40). Brains were dissected and post-fixed in cold Xgal fixative overnight at 4 °C. Brains were then washed and cryoprotected as described above and then labelled with Xgal enzyme. The reaction was monitored closely and stopped when background started to appear in control (*lacZ* negative) tissues. Following four washes in PBS, 30 μm cryosections from *Opn5*^{lacZ} mice were briefly post-fixed in 4% paraformaldehyde, counterstained with Nuclear Fast Red, dehydrated, and then imaged under standard transmitted brightfield using Zeiss AxioVision v4.9.1 SP2 software. For *Lepr-cre;Ai14;Opn5*^{lacZ} cryosections, the Nuclear Fast Red counterstain was not applied.

Cell size quantification (inWAT)

Haematoxylin-stained paraffin sectioned inWAT samples were imaged under 594 nm excitation through a rhodamine filter. Monochrome images were thresholded and adipocyte cell boundaries automatically detected using NIS Elements Advanced Research v.5.20.00 software (Nikon Instruments). Individual cells were demarcated as separate objects in a binary layer, filtered for circularity and size, and their area measured in μm^2. Approximately 500–1,000 cells were measured per field, with at least 20 fields per animal analysed, for a total of 10,000–20,000 cells per animal. Cell areas were binned into 100 μm^2 intervals and the frequency of total cells (percentage) charted for each interval.

M-FISH

M-FISH experiments were performed with fresh-frozen brain tissue. In brief, P21 male and female *Opn5*$^{cre/+}$*;Ai14* mice were euthanized and their brains rapidly dissected into cryo-embedding medium. Embedded brains were snap-frozen in liquid nitrogen, and 14 μm cryosections of the POA were obtained and processed for M-FISH using the RNAscope Fluorescent Multiplex Reagent Kit V1 (ACDBio). Probes against the following mRNAs were used: *Slc32a1*, *Slc17a6*, *Adcyap1*, *Bdnf* and *tdTomato*. In situ hybridization was performed as per the manufacturer's protocol for fresh frozen tissue. In brief, POA sections were pre-treated by serial immersion of the slides in 1× PBS, nuclease-free water, and 100% ethanol at room temperature for two minutes each. Probe hybridization was achieved by incubating sections in 40 μl of mRNA target probes for 2 h at 40 °C, followed by signal amplification using manufacturer-provided Amp1, Amp2, Amp3, and Amp4 reagents for 30, 15, 30, and 15 min respectively at 40 °C. Each incubation step was followed by two 2-min washes of manufacturer-provided wash buffer. Slides were mounted using Tris-buffered Fluoro-Gel mounting medium (Electron Microscopy Sciences).

M-FISH quantification

60× fields were acquired from *Opn5*$^{cre/+}$*;Ai14* POA regions from $n = 3$ mice. Before cell counting, negative control regions of interest (ROI) were acquired. Single cell images (715-μm^2 ROIs) of ependymal cells or dural cells were acquired to calculate background labelling for all 3 channels, which varied across experiments and probes. Using the nuclear marker channel (DAPI) and *tdTomato* (C2) probe, several 715μm^2 ROIs were acquired representing cells of interest. Then, puncta from C1 (*Slc32a1*, *Bdnf*, *Ptgds*) and C3 (*Slc17a6*, *Adcyap1* or *Trpm2*) for each ROI was calculated and the cell was assessed to be positive or negative for a marker. Cells were considered positive if the number of puncta was 1.5× above background for that section. A total of 109 cells from $n = 3$ mice were used for *Slc32a1* and *Slc17a6* assessment, a total of 92 cells from $n = 3$ mice for *Bdnf* and *Adcyap1* assessment, and a total of 92 cells from $n = 3$ mice were used for *Ptgds* and *Trpm2* assessment.

Serum lipids and thyroid hormones

Serum from P90-P120 *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ male and female mice were harvested and snap frozen. Lipid profiles (TG, PL, CHOL, NEFA) were obtained via standard colorimetric methods performed at the University of Cincinnati Mouse Metabolic Phenotyping Center (NIH 2U2C-DK059630-16). In brief, triglyceride quantification was performed by the GPO-PAP method (Randox), phospholipids by the choline oxidase-DAOS method (Wako Diagnostics), cholesterol by the Infinity cholesterol liquid stable reagent method (Thermo Scientific), and NEFAs by the ACS-ACOD method (Wako Diagnostics). Colorimetric measurements were obtained using a Synergy HT (BioTek) with Gen5 software. Serum measurements of free thyroxine (T4) and thyrotropin releasing hormone (TRH) were made using competitive ELISA and performed at the University of Massachusetts MMPC (NIH 5U2C-DK093000-07).

Western blotting

Western blots were performed using standard protocols. BAT from mice were dissected into 400 μl of modified RIPA lysis buffer and homogenized (Tissue Lyser II, Qiagen) using zirconium oxide beads (2.0 mm). After centrifugation and protein quantification (Pierce BCA Protein Assay Kit), 10 μg protein were loaded onto a 16% Novex Tris-Glycine protein gel and transferred to a PVDF (polyvinylidene difluoride) membrane, where bands were visualized by chemiluminescence. Antibodies used for western blotting include anti-UCP1 (Abcam, ab10983, 1:5,000 dilution) and anti-alpha tubulin (Abcam, ab4074, 1:5,000 dilution).

Quantitative RT–PCR

Intrascapular adipose depots were collected immediately following cold exposure assays. Snap frozen tissue was homogenized in TRI Reagent (Invitrogen) using RNase-free zirconium oxide beads (2.0 mm) in a TissueLyser II sample disrupter (Qiagen). Phase separation was accomplished via chloroform and RNA in the aqueous phase was precipitated with ethanol and column-purified via the GeneJET RNA purification kit (ThermoFisher Scientific K0732). Purified RNA was subsequently treated with RNase-free DNase I (ThermoFisher Scientific EN0521) and cDNA was synthesized using a Verso cDNA synthesis kit (ThermoFisher Scientific AB1453/B). Quantitative RT–PCR was performed with Radiant SYBR Green Lo-ROX qPCR mix (Alkali Scientific) in a ThermoFisher QuantStudio 6 & 7 Flex Real-Time PCR system. Primer information for quantitative PCR is included in Supplementary Table 2. Relative expression was calculated by the $\Delta\Delta C_T$ method using *Tbp* (TATA binding protein) as the normalizing gene. Statistical significance was calculated by a two-way ANOVA followed by Tukey post hoc analysis, using a *P*-value cutoff of 0.05.

Statistics and reproducibility

Statistical and image analyses were performed with MATLAB 2018a, NIS Elements Ar v.5.20.00, and ImageJ. Sample sizes for each experiment are reported in the Article or in the figures. The numbers of experimental repetitions were as follows: Fig. 1a, b, 12 times; Fig. 1c, d, 3 times; Fig. 1e–k, 3 times; Fig. 1l–x, 7 times; Fig. 2a–i, 5 times; Fig. 2j–u, 4 times; Fig. 3a–d, 5 times; Fig. 3e, twice; Fig. 3f–h, twice; Fig. 4a–k, 4 times; Extended Data Fig. 1a–k, 3 times; Extended Data Fig. 2a–c, 3 times; Extended Data Fig. 3a–l, 4 times; Extended Data Fig. 4a, 3 times, Extended Data Fig. 4b, twice (22 °C), once (4 °C). Extended Data Fig. 4c–e, 3 times; Extended Data Fig. 4f, g, 3 times; Extended Data Fig. 4h, i, once; Extended Data Fig. 4j–m, 4 times; Extended Data Fig. 4n–q, 3 times; Extended Data Fig. 4r, 3 times; Extended Data Fig. 4s, twice (22 °C), once (4 °C). Extended Data Fig. 4t–v, 3 times; Extended Data Fig. 5a-g, twice; Extended Data Fig. 6a–f, twice; Extended Data Fig. 6g–i, 3 times; Extended Data Fig. 6j–l, twice; Extended Data Fig. 6m, 3 times; Extended Data Fig. 7a–g, 3 times; Extended Data Fig. 8a–i, 6 times; Extended Data Fig. 9c, 3 times; Extended Data Fig. 9d, twice; Extended Data Fig. 9e–g, 3 times; Extended Data Fig. 9h, i, twice; Extended Data Fig. 9j–m, 3 times; Extended Data Fig. 10a–e, 3 times.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Source data are provided with this paper. All other relevant data are available from the corresponding authors upon request. Source data are provided with this paper.

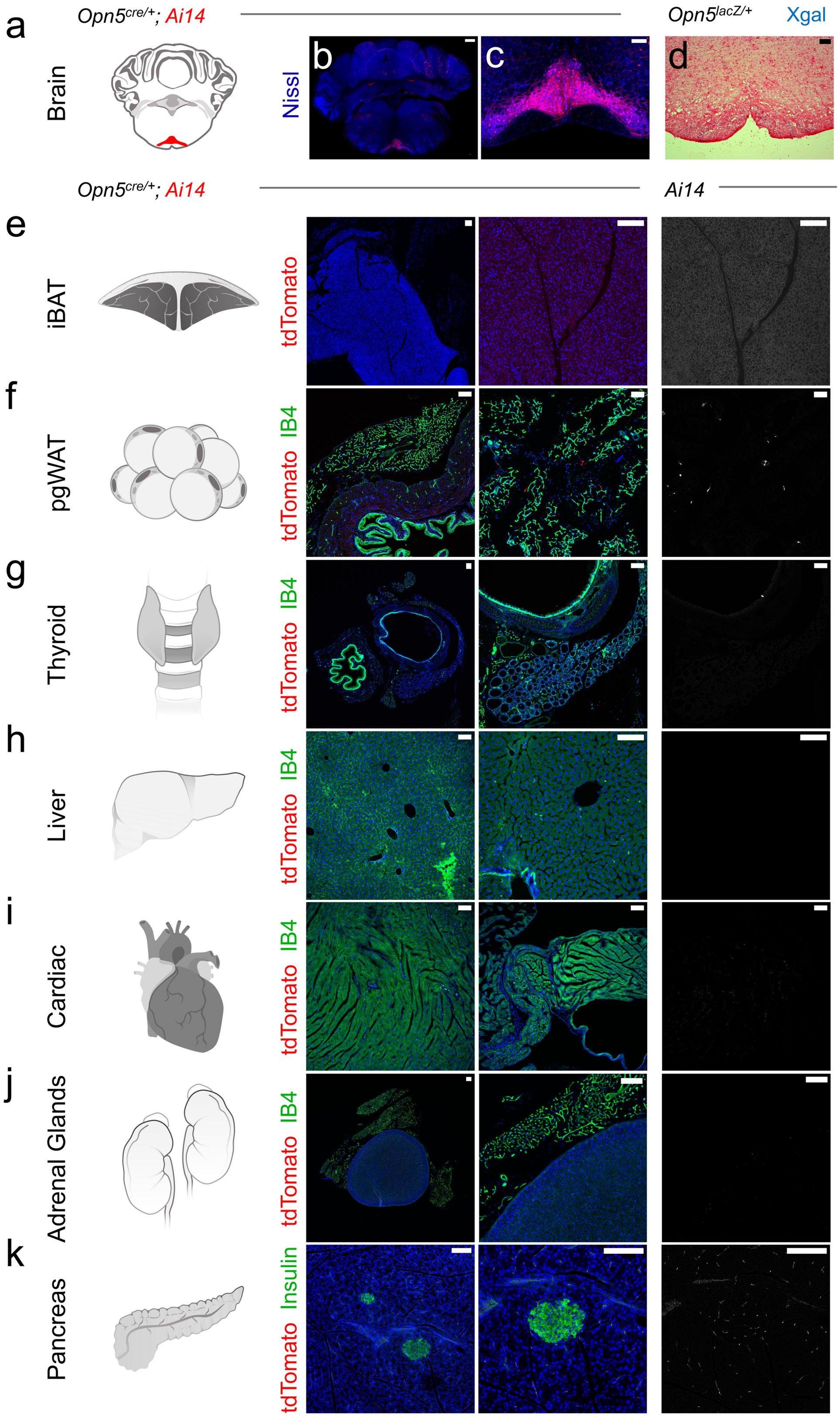
30. Holt, A. L., Vahidinia, S., Gagnon, Y. L., Morse, D. E. & Sweeney, A. M. Photosymbiotic giant clams are transformers of solar flux. *J. R. Soc. Interface* **11**, 20140678 (2014).

Acknowledgements We thank P. Speeg for mouse colony management; Y. Chen and Y.-C. Hu of the CCHMC Transgenic Animal and Genome Editing Core Facility for mouse line development; M. Kofron of the CCHMC Confocal Imaging Core Facility for assistance; M. Talley for consultation on M-FISH; T. Nakamura, V. Borra and A. Vonberg for assistance with metabolic experimentation; A. Ahmed for technical assistance; J. Lighton and B. Joos of Sable Systems International for assistance with indirect calorimetry; and T. Delehanty of TSE Systems for assistance with temperature telemetry. This work was supported by NIGMS 5T32GM063483 (University of Cincinnati MSTP), NEI R01 s EY027711 and EY027077 (to R.A.L.); NIGMS R01 GM124246 (to E.D.B.), NEI R01 EY026921 (to R.N.V.G.), NEI P30 EY001730 (to the Vision Research Core at the University of Washington); NIDDK P30 DK089503 (to R.J.S. and the Michigan Nutrition Obesity Research Center); the Mark J. Daily, MD Research Fund (to the University of Washington); and unrestricted grants to the University of Washington Department of Ophthalmology from Research to Prevent Blindness. This work was also supported by a Packard Foundation fellowship (to A.S.); American Heart Association grant 18CDA34080527 (to J.S.G.); and funds from the Goldman Chair of the Abrahamson Pediatric Eye Institute at CCHMC.

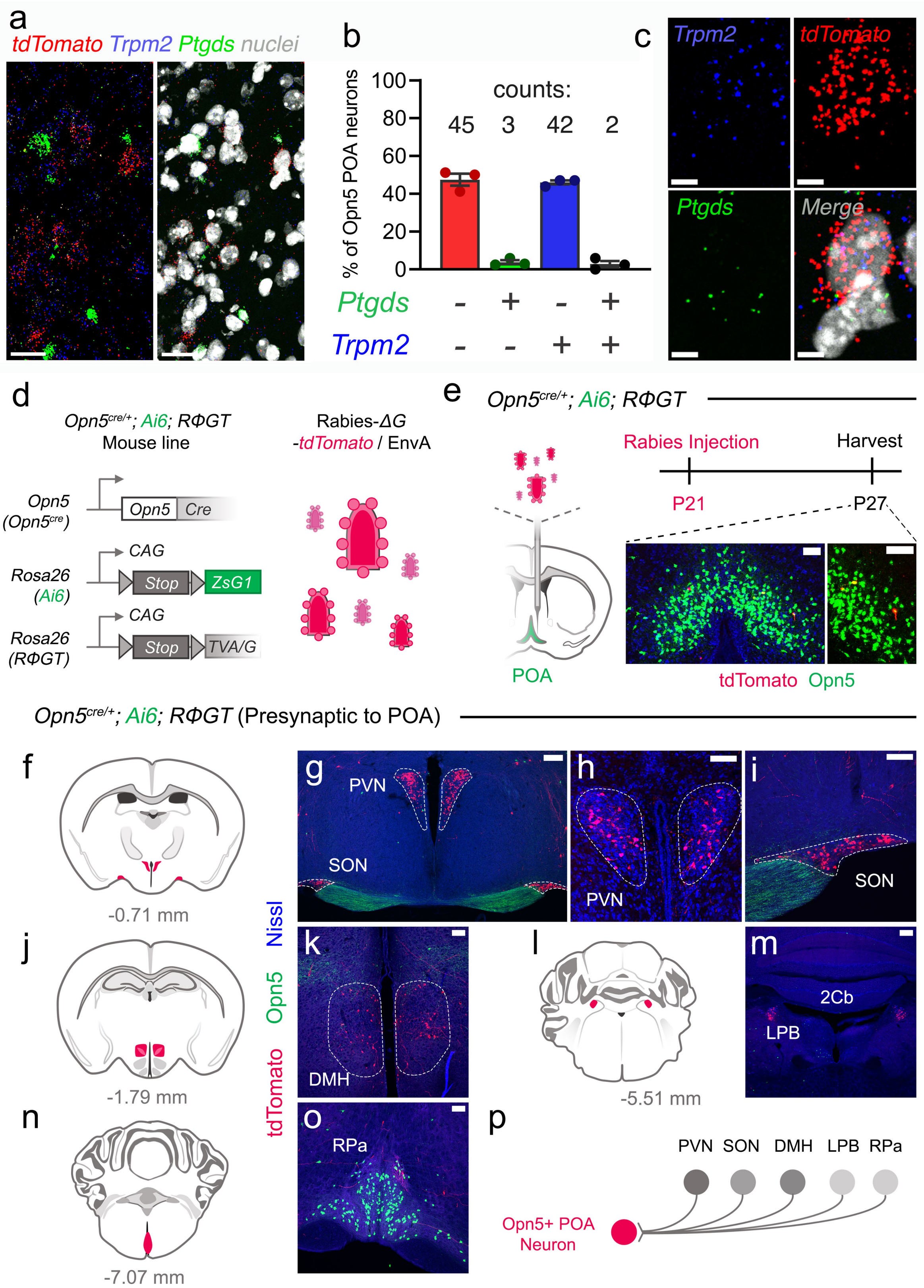
Author contributions S.V., K.X.Z. and R.A.L. conceived and directed the study. K.X.Z. and S.D. performed imaging experiments and quantitative analysis. S.D. performed M-FISH experiments. K.X.Z. performed stereotaxic and telemetry surgeries, and chemogenetic experiments. K.X.Z., S.D., B.A.U., S.V. and G.N. performed cold exposure experiments. C.D.L. and R.M. performed western blotting. K.X.Z. performed viral tracing studies. K.X.Z. and K.D.G. performed CAMPER experiments. K.X.Z. and B.A.U. performed locomotion experiments. S.K. and R.J.S. performed indirect calorimetry experiments. A.H.J. and A.S. provided expertise in radiometry and performed radiometric experiments with K.X.Z. and B.A.U. N.T.P. and M.B. designed and constructed custom electronic equipment. A.N.S. performed Xgal labelling. D.T. assisted with neuroanatomical studies. E.D.B., R.N.V.G., C.G., J.S.G. and R.J.S. consulted on experimental design and reviewed the manuscript. K.X.Z. and S.D. analysed data. K.X.Z., S.D. and R.A.L. wrote the manuscript. R.A.L. provided coordinating project leadership.

Competing interests R.J.S. receives research support from Novo Nordisk, Zafgen, Kallyope, Pfizer, and Ironwood Pharmaceuticals.

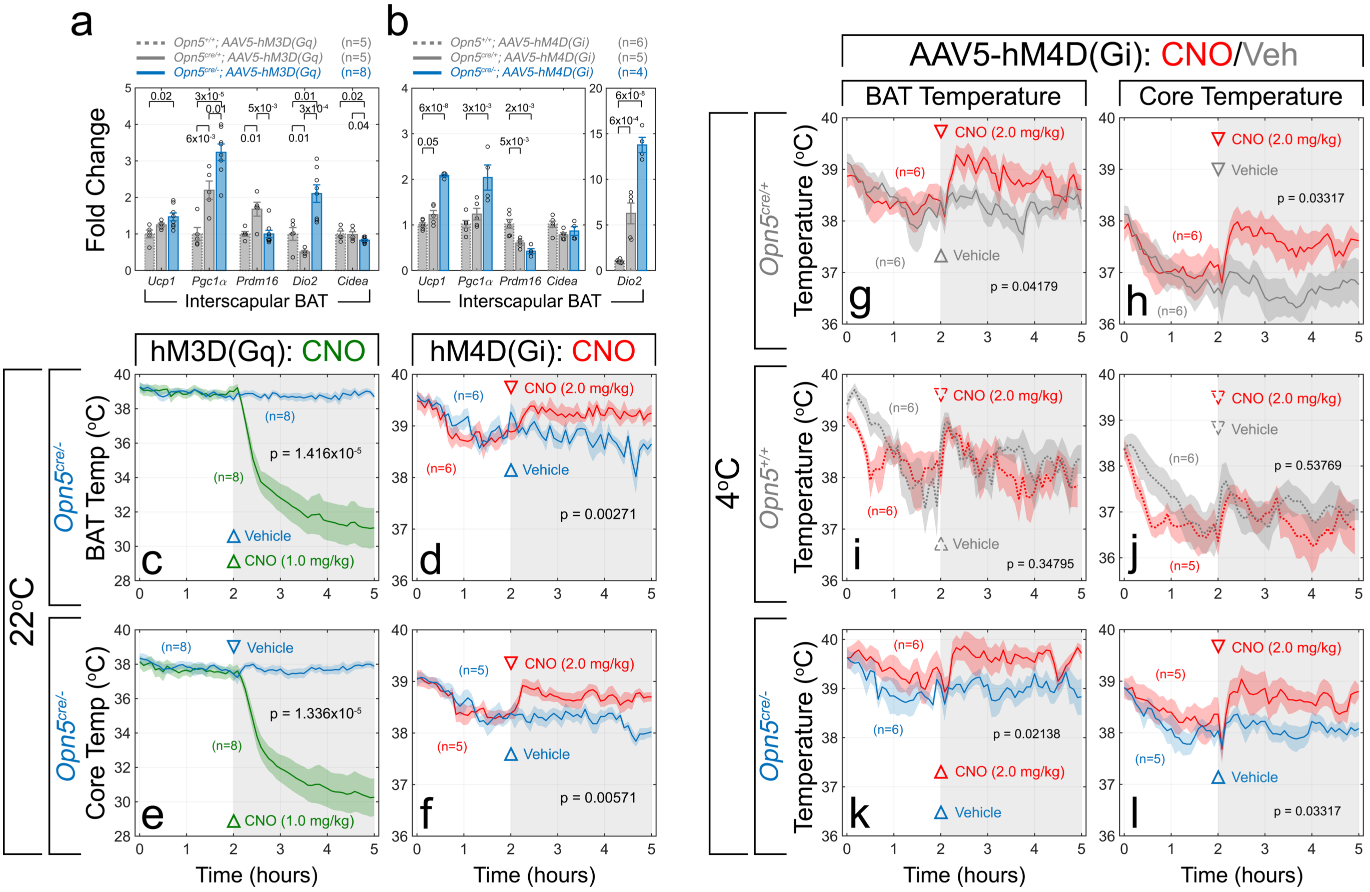
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2683-0.
Correspondence and requests for materials should be addressed to R.A.L.
Peer review information Peer reviewer reports are available.
Reprints and permissions information is available at http://www.nature.com/reprints.



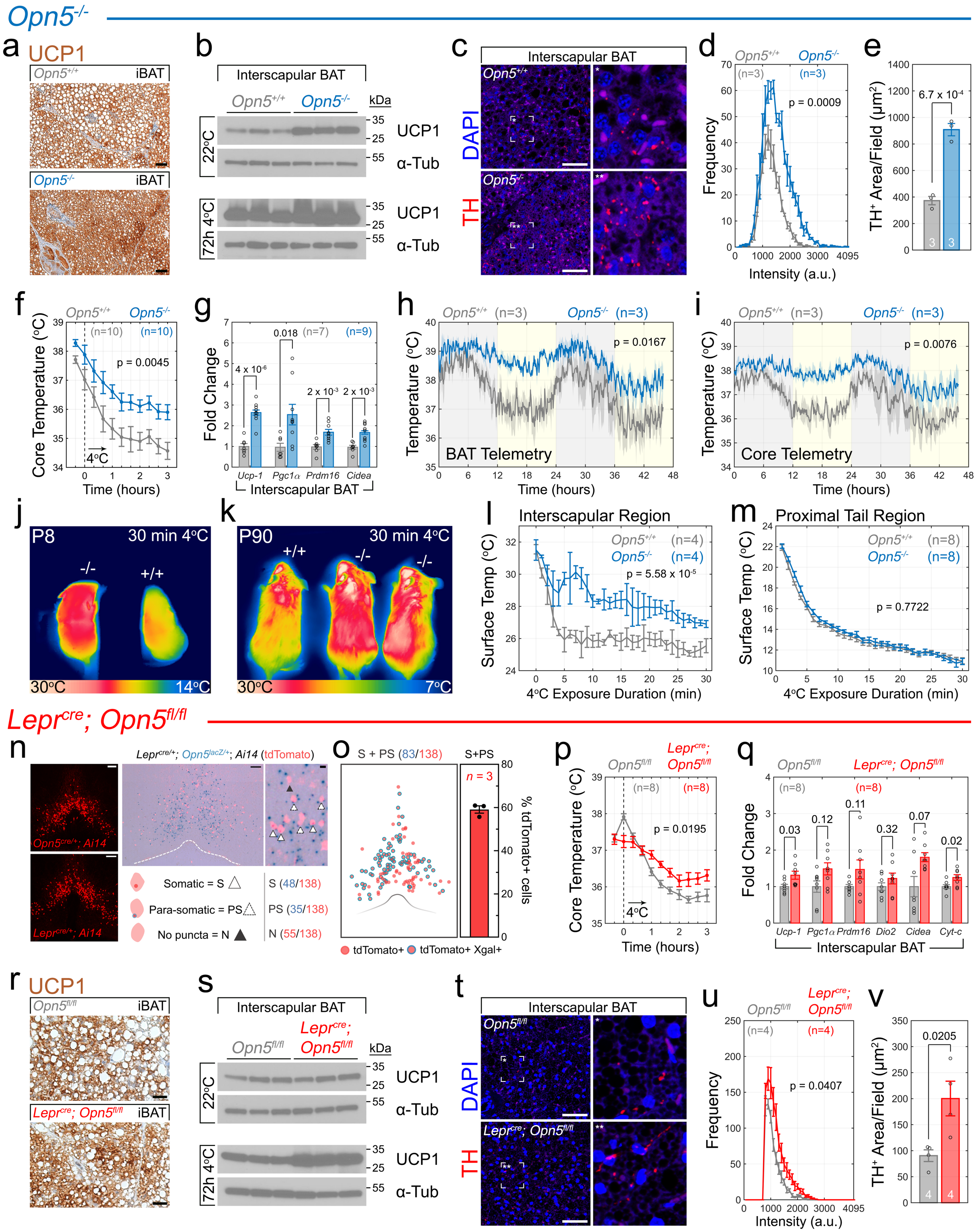
Extended Data Fig. 1 | *Opn5* lineage survey across the CNS and thermogenic organs. **a**–**c**, Brain atlas representation (**a**) and coronal brain sections (**b**, **c**) of P21 *Opn5*$^{cre/+}$;*Ai14* mouse highlighting tdTomato expression (red) in the raphe pallidus (RPa). **d**, Coronal brain section from P10 *Opn5*$^{lacZ/+}$ mouse showing that the RPa is negative for Xgal labelling. **e**–**k**, Representative confocal images from IB4 labelled (green) P35 *Opn5*$^{cre/+}$;*Ai14* (expressing tdTomato, red) tissues across the organism. iBAT (**e**), perigonadal white adipose tissue (pgWAT) (**f**), thyroid gland (**g**), liver (**h**), cardiac muscle (**i**), adrenal glands (**j**), and pancreas (**k**). Scale bars, 100 µm (**e**–**k**), 150 µm (**c**, **d**), 500 µm (**b**).



Extended Data Fig. 2 | M-FISH and presynaptic tracing of *Opn5* POA neurons. **a**, **b**, Representative images of *Opn5*$^{cre/+}$;*Ai14* POA neurons probed for *Ptgds* (green), *tdTomato* (red), *Trpm2* (blue) and labelled with DAPI for nuclei (greyscale) (**a**) with corresponding quantification of overlap ($n = 3$ mice; 92 cells) (**b**). **c**, Representative images of *Opn5*$^{cre/+}$;*Ai14* cells (red) also positive for *Trpm2* (blue) but with *Ptgds* labelling (green) that is below the background labelling threshold. Scale bars, 5 µm (**c**), 25 µm (**a**). **d**, Schematic of the mouse genetics used for rabies viral tracing. **e**, Experimental timeline for POA-tracing, and primary infected neurons (yellow). **f**–**o**, Traced neurons (red) located in the paraventricular nucleus (PVN) (**f**–**h**), supraoptic nucleus (SON) (**f**, **g**, **i**), dorsomedial hypothalamus (DMH) (**j**, **k**), lateral parabrachial (LBP) (**l**, **m**), and raphe pallidus (RPa) (**n**, **o**). Green regions in **g** and **i** are optic tracts with axons from *Opn5* retinal ganglion cells. **p**, Schematic representation of nuclei presynaptic to *Opn5* POA neurons. Data in **b** are mean ± s.e.m. Scale bars, 5 µm (**c**), 25 µm (**a**), 75 µm (**e**), 100 µm (**h**, **i**, **k**, **o**) or 200 µm (**g**, **m**). 2Cb, lobule 2 of cerebellar vermis.

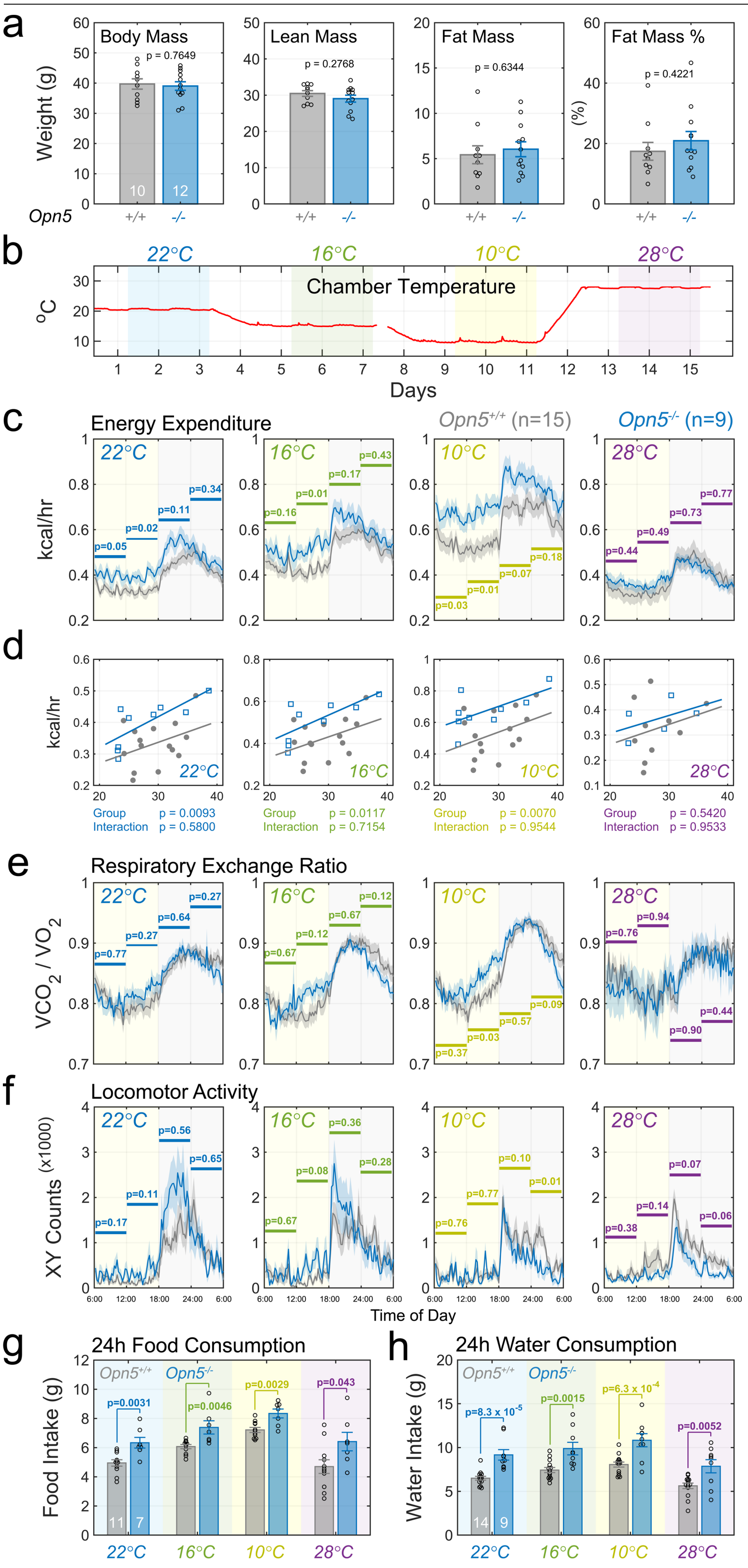


Extended Data Fig. 3 | Thermoregulation by *Opn5* POA neurons is not context-dependent. a, **b**, qPCR of thermogenesis genes in iBAT 6 h after CNO induction in mice with viral-mediated expression of stimulatory hM3D(G_q) DREADD (**a**) or inhibitory hM4D(G_i) DREADD (**b**) in the POA. (**a**) *Opn5*$^{+/+}$ (*n* = 5), *Opn5*$^{cre/+}$ (*n* = 5), and *Opn5*$^{cre/-}$ (*n* = 8). (**b**) *Opn5*$^{+/+}$ (*n* = 6), *Opn5*$^{cre/+}$ (*n* = 5), and *Opn5*$^{cre/-}$ (*n* = 4). *P* values are indicated above the bars. **c**–**f**, Similar to Fig. 2, *Opn5*$^{cre/-}$ POA was injected with AAV5-hM3D(G_q) DREADD (**c**, **e**; *n* = 8 mice per condition) or AAV5-hM4D(G_i) DREADD (**d**, **f**; *n* = 6 mice per condition). Telemetric BAT and core recordings after intraperitoneal administration of CNO or vehicle (open arrowheads) at the 2 h mark. **g**–**l**, *Opn5*$^{+/+}$, *Opn5*$^{cre/+}$, *Opn5*$^{cre/-}$ (*n* = 6 per genotype and condition) mice were injected with AAV5-hM4D(G_i) DREADD, and followed by chemogenetic manipulations as previously described, but at 4 °C ambient temperature. All data are mean ± s.e.m. *P* values are from (**a**, **b**) ANOVA with Tukey post hoc analysis (**a**, **b**), or one-way repeated measures ANOVA (**c**, **l**).

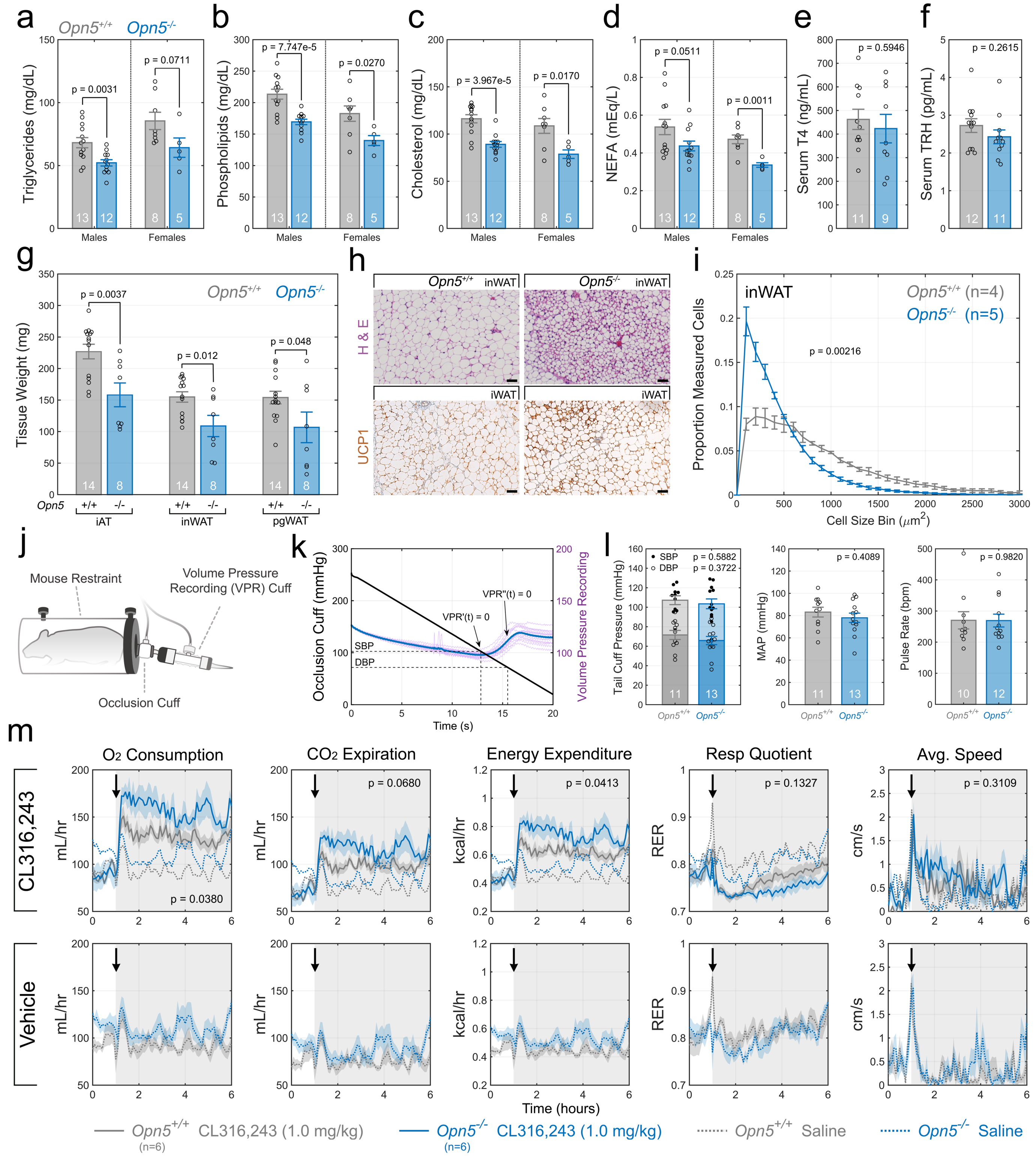


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | *Opn5* loss-of-function exaggerates BAT thermogenesis. **a**, Immunohistochemistry for UCP1 protein in iBAT from *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice. **b**, UCP1 immunoblots for iBAT comparing ambient temperature (22 °C) and 72 h 4 °C exposure for *Opn5*$^{+/+}$ ($n = 3$) and *Opn5*$^{-/-}$ ($n = 3$) mice. **c**–**e**, Representative immunofluorescence of TH^{+} innervation of iBAT (**c**) used for quantification in **d** and **e**. **f**, Core temperature assessment (rectal) of *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice during 3 h cold exposure. **g**, qPCR of thermogenesis genes (*Ucp1*, *Pgc1a*, *Prdm16* and *Cidea*) in iBAT from mice in **f**. **h**, **i**, Forty-eight hour assessment of body temperature rhythms in *Opn5*$^{+/+}$ ($n = 3$ mice) and *Opn5*$^{-/-}$ ($n = 3$ mice) mice using telemetry sensors in iBAT (**h**) and core (**i**) under 12 h/12 h light/dark lighting conditions. **j**, **k**, Infrared thermography of P8 (**j**) and P90 (**k**) *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice following 30 min cold challenge. **l**, **m**, Quantification of thermographic images focused on interscapular region (**l**), and tail (**m**). **n**, Representative POA images from *Opn5*$^{cre/+}$*;Ai14* and *Lepr*$^{cre/+}$*; Ai14* mice, plus *Lepr*$^{cre/+}$*;Ai14* colocalization with *Opn5*$^{lacZ/+}$ expression (Xgal). **o**, Quantification of overlap in **n**. **p**, Core temperature assessment (rectal) of control (*Opn5*$^{fl/fl}$) and *Lepr*^{cre}*; Opn5*$^{fl/fl}$ mice during 3 h cold challenge. **q**, qPCR of thermogenesis genes in iBAT from mice in **p**. **r**, Immunohistochemistry for UCP1 protein in iBAT from *Opn5*$^{fl/fl}$ and *Lepr*^{cre}*; Opn5*$^{fl/fl}$ mice. **s**, UCP1 immunoblots for iBAT comparing ambient temperature (22 °C) and 72 h 4 °C exposure for *Opn5*$^{fl/fl}$ ($n = 3$) and *Lepr*^{cre}*;Opn5*$^{fl/fl}$ ($n = 3$) mice. **t**–**v**, Representative immunofluorescence of TH^{+} innervation of iBAT (**t**) used for quantification in **u** and **v**. Scale bars, 50 μm (**a**, **c**, **r**, **t**), 100 μm (**n**). Data are mean ± s.e.m. *P* values are from one-way repeated measures ANOVA (**d**, **f**, **h**, **i**, **l**, **m**, **p**, **u**), ANOVA with Tukey post hoc analysis (**g**, **q**) or two-tailed Student's *t*-test (**e**, **v**).

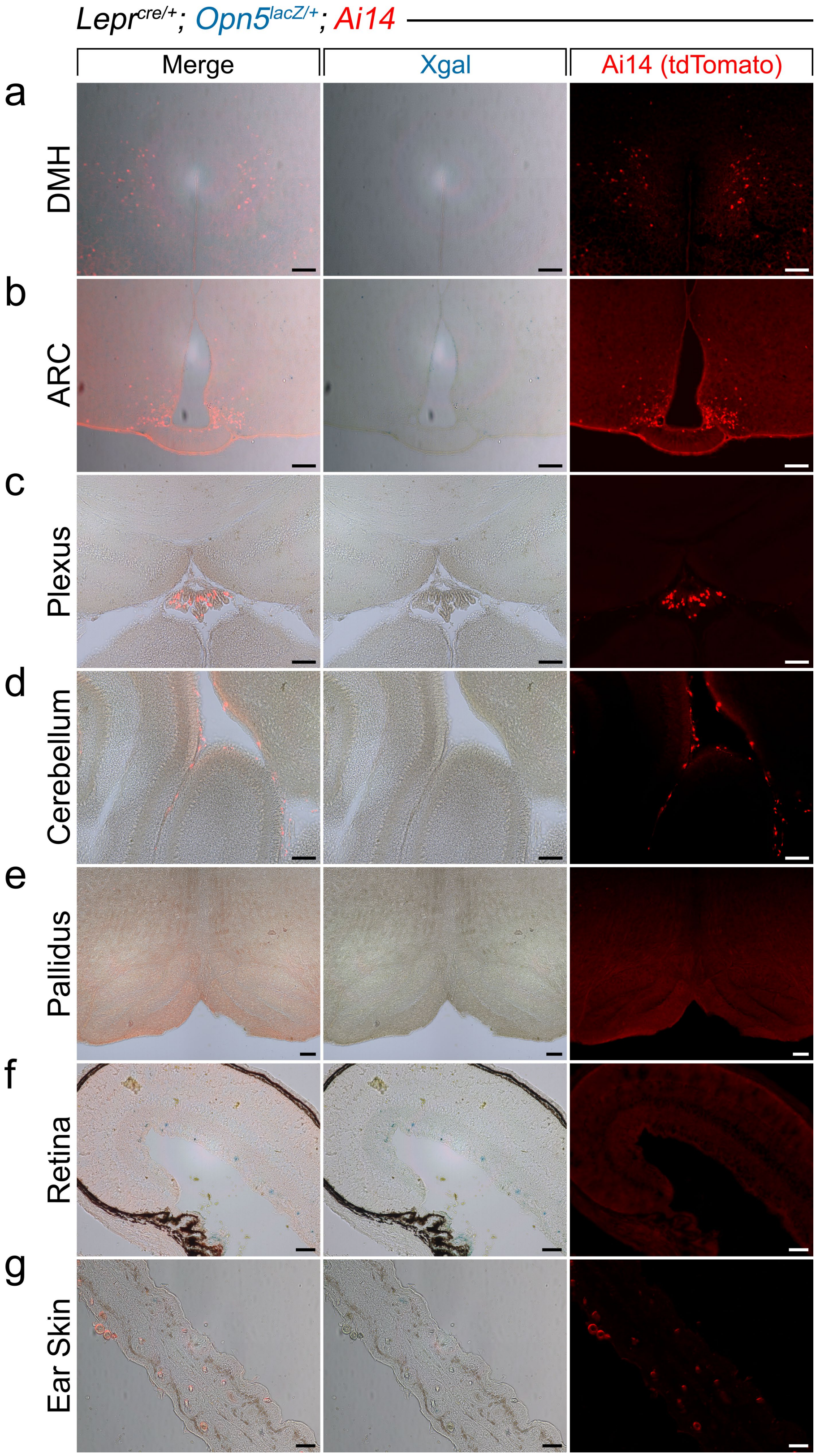


Extended Data Fig. 5 | *Opn5* null mice have altered energy homeostasis. a, Body mass, body composition (lean mass/fat mass), and fat mass as a percentage of body mass (fat mass percentage) comparison between *Opn5*$^{+/+}$ (*n* = 10) and *Opn5*$^{-/-}$ (*n* = 12) mice. **b**, Schematic describing ambient temperature changes throughout experiment and the duration of measurement intervals. **c**, Indirect calorimetry (TSE Systems, PhenoMaster Cages) measurements of energy expenditure in adult *Opn5*$^{+/+}$ (grey trace, *n* = 15) and *Opn5*$^{-/-}$ (blue trace, *n* = 9) mice at ambient temperatures of 22 °C, 16 °C, 10 °C and 28 °C. **d**, Mass–energy relationships of data in **c** represented as generalized linear models. **e**, Respiratory exchange ratio (RER = VCO_2/VO_2) obtained from the same mice. **f**, Spontaneous locomotor activity (XY) monitoring was performed via infrared beam breaks. **g**, Twenty-four-hour average food consumption from *Opn5*$^{+/+}$ (grey bars, *n* = 11) and *Opn5*$^{-/-}$ (blue bars, *n* = 7) mice at each ambient temperature. Mice exhibiting 'food grinding' behaviour were excluded from the analysis. **h**, Twenty-four-hour average water consumption from *Opn5*$^{+/+}$ (grey bars, *n* = 14) and *Opn5*$^{-/-}$ (blue bars, *n* = 9) mice at each temperature. *P* values are from two-tailed Student's *t*-tests (**a**), one-way repeated measures ANOVA across 6-h time interval (**c**, **e**, **f**), two-way ANCOVA with body mass as covariate (**d**), or ANOVA with Holm–Sîdak corrected multiple comparisons (**g**, **h**). Data in **c**, **e**, **f** show a 24h period of mean ± s.e.m. data for both genotypes during lights on (06:00–18:00, yellow shaded region) followed by lights off (18:00–06:00, grey shaded region). Data in **a**, **g** and **h** are represented as mean ± s.e.m.

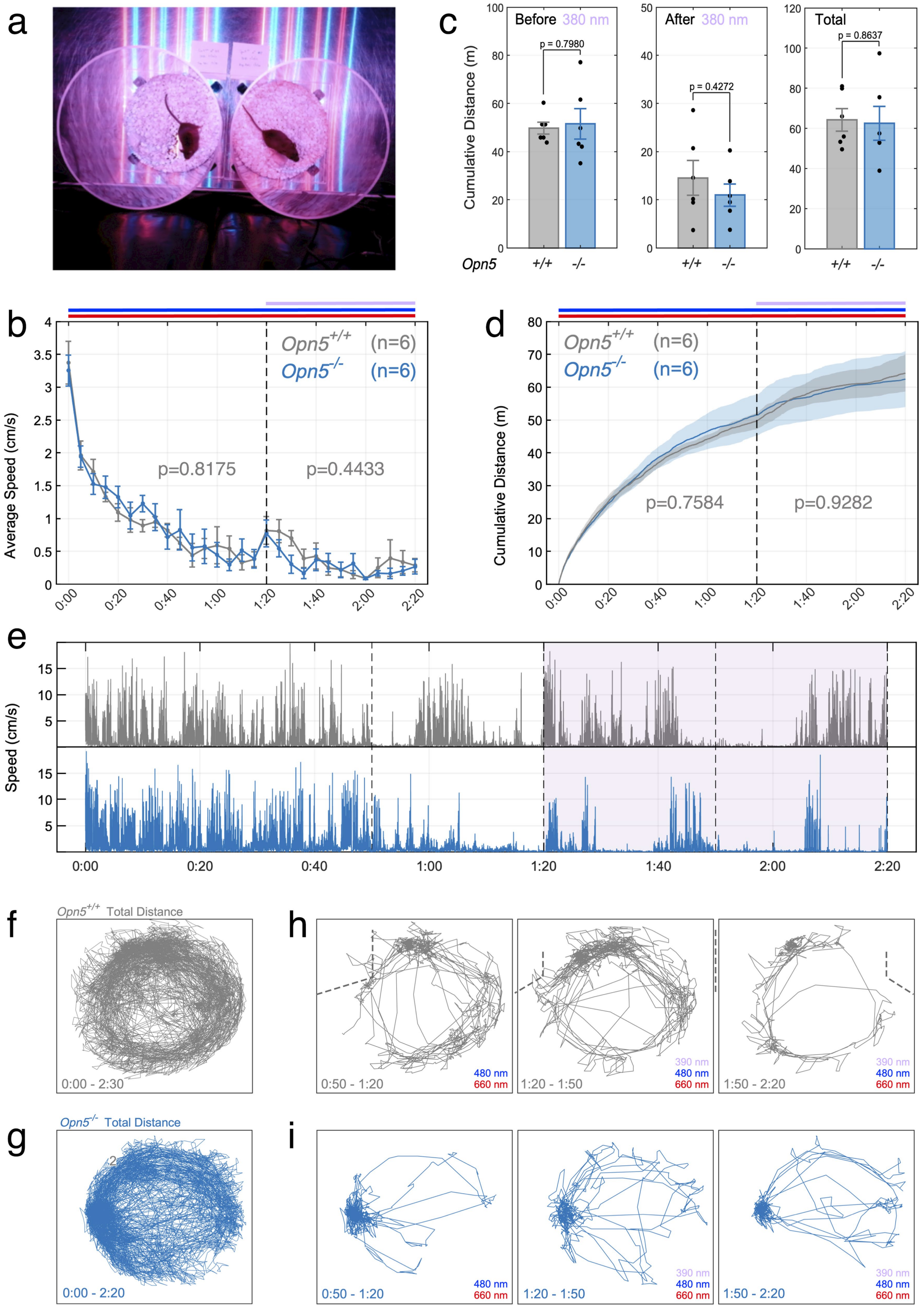


Extended Data Fig. 6 | See next page for caption.

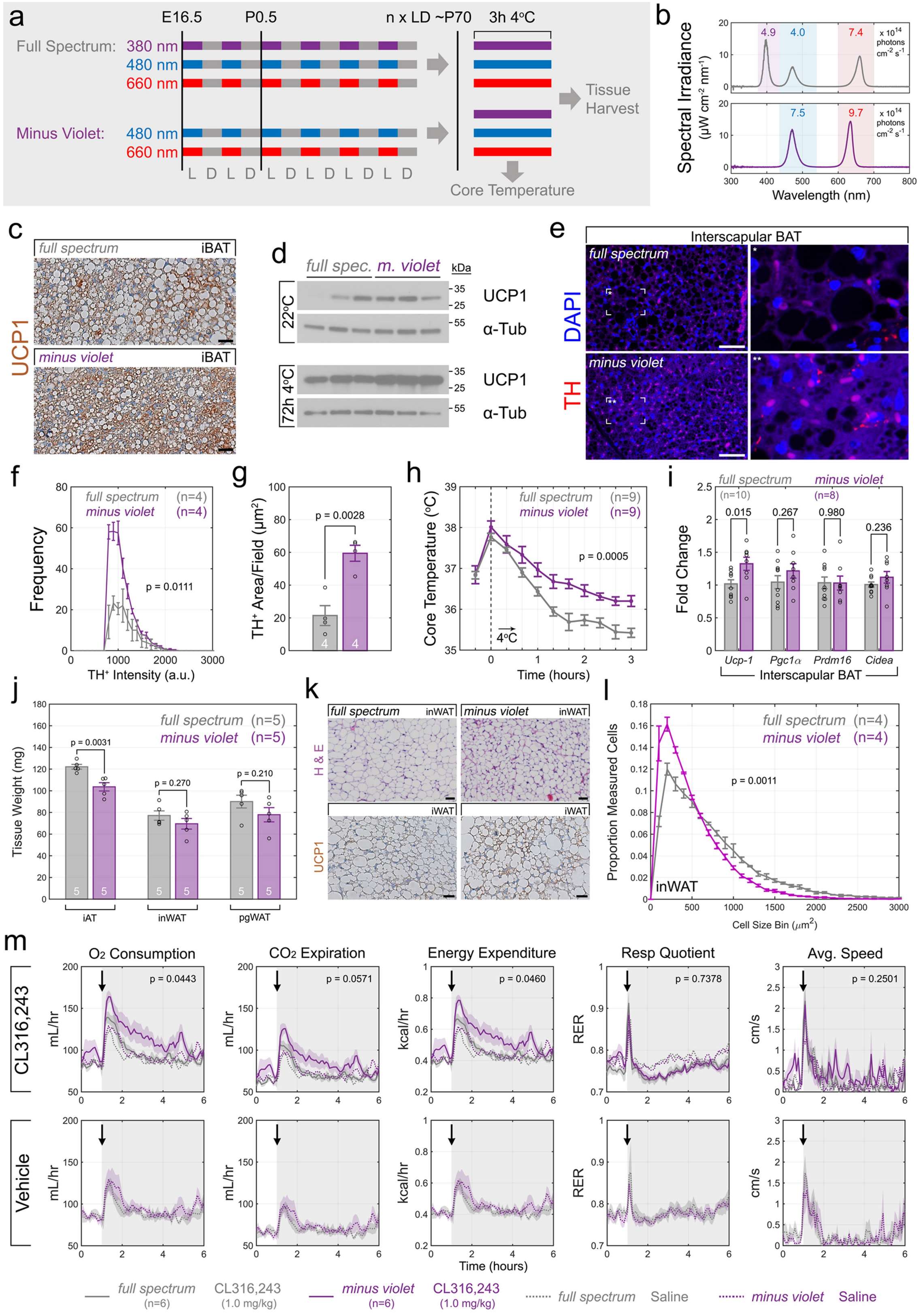
Extended Data Fig. 6 | OPN5 regulates thermogenesis and lipid metabolism, but not thyroid and cardiovascular activity. **a**–**d**, Serum lipid quantifications from male ($n = 13$ *Opn5*$^{+/+}$, $n = 12$ *Opn5*$^{-/-}$) and female ($n = 8$ *Opn5*$^{+/+}$, $n = 5$ *Opn5*$^{-/-}$) mice for triglycerides (**a**), phospholipids (**b**), cholesterol (**c**), and non-esterified fatty acids (NEFA) (**d**). **e**, Serum thyroxine (T4) from male *Opn5*$^{+/+}$ ($n = 11$) and *Opn5*$^{-/-}$ ($n = 9$) mice. **f**, Serum thyrotropin-releasing hormone (TRH) from male *Opn5*$^{+/+}$ ($n = 12$) and *Opn5*$^{-/-}$ ($n = 11$) mice. **g**, Adipose depot weight (mg) comparison between male *Opn5*$^{+/+}$ ($n = 14$) and *Opn5*$^{-/-}$ ($n = 8$) mice. inWAT, inguinal white adipose tissue. **h**, Representative images highlighting inWAT cell size (H&E) and iWAT UCP1 (IHC) from *Opn5*$^{+/+}$ and *Opn5*$^{-/-}$ mice. Scale bars, 50 µm. **i**, Quantification of inWAT cell size for *Opn5*$^{+/+}$ ($n = 4$) and *Opn5*$^{-/-}$ ($n = 5$) mice. **j**, Schematic representation of mouse blood pressure recording system. Animals are movement-restricted in a mouse restraint and the tail is fitted proximally with an occlusion cuff and distally with a volume pressure recording (VPR) cuff. **k**, Example trial from tail blood pressure recording. Data are represented as line graphs for occlusion cuff pressure (mmHg; left *y*-axis) and VPR cuff pressure (mmHg; right *y*-axis). **l**, Quantification of blood pressure (SBP, systolic blood pressure; DBP, diastolic blood pressure), mean arterial pressure (MAP), and pulse rate (bpm) from *Opn5*$^{+/+}$ ($n = 10$-11) and *Opn5*$^{-/-}$ ($n = 12$-13) mice. **m**, Indirect calorimetry and locomotion from *Opn5*$^{+/+}$ ($n = 6$, grey trace) and *Opn5*$^{-/-}$ ($n = 6$, blue trace) mice treated with 1.0 mg/kg β_3 adrenergic receptor agonist CL-316,243 (solid line) or vehicle control (saline, dotted line). Intraperitoneal injection of agonist or saline was performed at the 1 h time point (indicated by arrow). All data are mean ± s.e.m. *p* values are from (**a**–**g**, **l**) two-tailed Student's *t*-test, (**i**, **m**) 1-way repeated measures ANOVA.



Extended Data Fig. 7 | Overlap of *Lepr* and *Opn5* expression is limited to the POA. **a**–**g**, *Lepr*-lineage and *Opn5* expression survey across multiple tissues ($n = 3$ mice). Representative images of tdTomato (*Lepr*^{cre}*; Ai14*) and Xgal (*Opn5*$^{lacZ/+}$) domains from the dorsomedial hypothalamus (**a**), arcuate nucleus (ARC) (**b**), choroid plexus (**c**), cerebellum (**d**), raphe pallidus (**e**), retina (**f**), and ear skin (**g**). Scale bars, 100 µm (**a**–**e**), 50 µm (**f**, **g**).

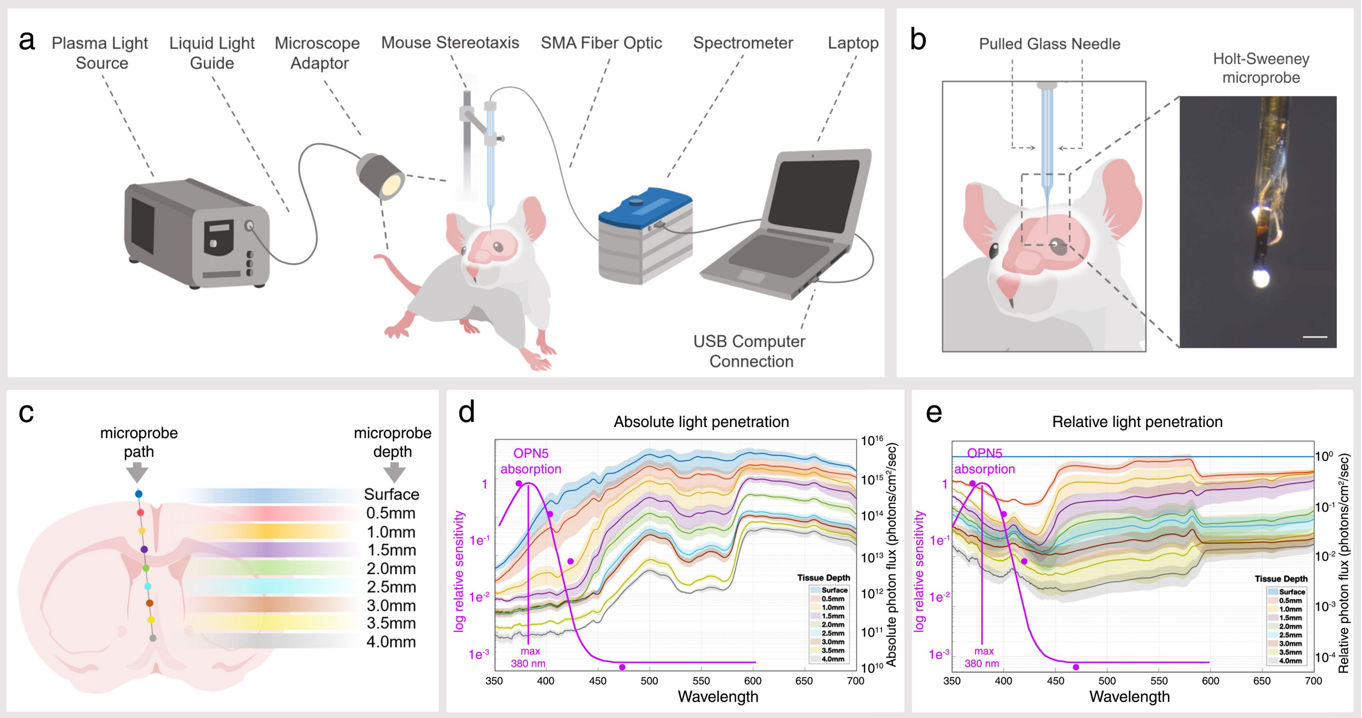


Extended Data Fig. 8 | Violet light does not change locomotor behaviour in cold exposed mice. a, Photograph of experimental setup in 4 °C. **b**, Average speed in cm/s of 2-month-old male *Opn5*$^{+/+}$ (grey trace; $n = 6$) and *Opn5*$^{-/-}$ (blue trace; $n = 6$) mice binned in 5 min intervals. Violet 380 nm LEDs were switched on after 80 min (1:20 mark). **c**, Cumulative distance in meters travelled by *Opn5*$^{+/+}$ ($n = 6$) and *Opn5*$^{-/-}$ ($n = 6$) mice before and after violet supplementation, along with total cumulative distance. **d**, Total cumulative distance plotted across time. **e**, Absolute speed in cm/s of a representative pair ($n = 1$ *Opn5*$^{+/+}$ and $n = 1$ *Opn5*$^{-/-}$) of mice. **f**, **g**, Representative mouse locomotion trace (centroid-based motion tracking) of the *Opn5*$^{+/+}$ (**f**) and *Opn5*$^{-/-}$ (**g**) mouse from (**e**). **h**, **i**, Selective locomotion traces in 30 min bins ranging from 0:50 – 1:20, 1:20 – 1:50, and 1:50 – 2:20, for the *Opn5*$^{+/+}$ (**h**) and *Opn5*$^{-/-}$ (**i**) experimental pair of mice from (**e**). *P* values are from (**b**, **d**) 1-way repeated measures ANOVA, and (log *p* value graphs from **b** and **d**), (**c**) two-tailed Student's *t*-test. Data in (**b**–**d**) are represented as mean ± s.e.m.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Violet light deprivation alters BAT innervation and sensitivity to sympathetic nervous system input. a, Lighting protocol used to generate 'full spectrum' and 'minus violet' mice. **b**, Spectral quality of lighting used in 'full spectrum' (top) and 'minus violet' (bottom) housing. Colored boxes indicate wavelength bounds used to estimate flux (photons $cm^{-2}s^{-1}$). **c**, UCP1 IHC of 'full spectrum' (top) and 'minus violet' (bottom) mice. **d**, Immunoblots of UCP1 at baseline (22 °C) and following 72 h cold adaptation (72h 4 °C) between 'full spectrum' ($n = 3$) and 'minus violet' ($n = 3$) mice. **e**–**g**, Representative images (**e**) of TH+ (tyrosine hydroxylase) innervation of BAT used for quantification represented in (**f**) and (**g**). **h**, Core temperature assessment (rectal) of 'full spectrum' and 'minus violet' mice during a 3h cold challenge. **i**, QPCR of thermogenesis genes (*Ucp1, Pgc1α, Prdm16, Cidea*) in iBAT from the mice used in (**h**). **j**, Adipose depot weight (mg) comparison between 'full spectrum' ($n = 5$) and 'minus violet' ($n = 5$) mice. **k**, Representative images highlighting inWAT cell size (haematoxylin and eosin) and iWAT UCP1 (IHC). **l**, Quantification of inWAT cell size H&E images for 'full spectrum' ($n = 4$) and 'minus violet' ($n = 4$) groups. **m**, Indirect calorimetry from 'full spectrum' ($n = 6$, grey trace) and 'minus violet' ($n = 6$, purple trace) mice treated with 1.0 mg kg^{-1} β_3 adrenergic receptor agonist CL-316,243 (solid line) or vehicle (saline, dotted line). Administration of agonist or saline was performed at the 1 h time point (indicated by arrow). Data are mean ± s.e.m. P values are from one-way repeated measures ANOVA (**f**, **h**, **l**, **m**), ANOVA with Tukey post hoc analysis (**i**), or two-tailed Student's t-test (**g**, **j**). Scale bars, 50 µm.



Extended Data Fig. 10 | Measurement of photon flux within the POA. **a**, Schematic of experimental setup for measuring intra-cranial photon flux as described in Methods. **b**, Holt–Sweeney microprobe consisting of a pulled optic fibre with an attached transparent spherical diffusing tip. Scale bar, 100 µm. **c**, Measurement depths and probe path within cranium. **d**, Absolute photon flux within mouse cranium with OPN5 action spectrum superimposed (adapted from[2] with data points from[4]). Top blue trace represents surface flux and, at the λ_{max} of OPN5, is about 3.4×10^{13} photons $cm^{-2} s^{-1}$. At the maximum 4.0 mm depth (grey trace), the flux at the λ_{max} of OPN5 is approximately 9.5×10^{10} photons $cm^{-2} s^{-1}$. **e**, Relative photon flux normalized to surface measurements. Each trace is expressed as mean ± s.e.m. from $n = 3$ mice.

nature research

Corresponding author(s): Richard A. Lang

Last updated by author(s): Jul 5, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NIS Elements C v5.20.02 for confocal imaging; FLIR Tools Desktop 5.13.18031.2002 for IR imaging; Zeiss AxioVision 4.9.1 SP2 for brightfield imaging (Xgal); Applied Biosystems Quantstudio 6 and 7 Flex for RT-PCR; BIOPAC AcqKnowledge 5.0 for temperature telemetry; TSE-Phenomaster for indirect calorimetry; Ocean Insight OceanView 1.6.5 for intratissue radiometry; Sable Systems International MetaScreen v2.3.15.11 for indirect calorimetry; Kent Scientific Corporation CODA Data Acquisition Software 4.1 for blood pressure measurements.
Data analysis	NIS Elements Ar v5.20.00 and ImageJ 1.52p (FIJI) for image analysis; Sable Systems International ExpeData v1.9.27 for indirect calorimetry data export; MATLAB 2018a for data processing, statistical analysis, and figure generation.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Source data in Excel and MATLAB format for all experiments in this study are available under "Source Data" in this article. All other relevant data are available from the corresponding authors upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Statistical methods were used to predetermine sample size using the sampsizepwr() function in MATLAB 2018a with a cutoff of 0.80 for indirect calorimetry experiments. For experiments without predetermination, sample sizes were chosen on the basis of prior experience and published standards in the field, which are cited in the references section (PMIDs: 27616062, 27562954, 29298426, 27147656, 32079648). Sample sizes are indicated for each experiment in the manuscript.
Data exclusions	No data were excluded.
Replication	All experiments have been successfully repeated with similar results at least once. All data presented in figures represent the results from at least two or more independent experiments. Information on experimental repetitions for each figure is included in the "Statistics and reproducibility" section of the article.
Randomization	Primary grouping of mice were based on genotype and lighting condition (minus violet vs. full spectrum) and thus no randomization was required for this study.
Blinding	Investigator and other authors were blinded to genotype and lighting condition for core temperature assessments, fat dissections, western blotting, imaging of adipocyte size quantification, and general confocal imaging (viral tracing, M-FISH, TH immunohistochemistry). Blinding was not performed on indirect calorimetry experiments, infrared thermography, RT-PCR, indwelling temperature telemetry, intracellular cAMP imaging, noninvasive blood pressure measurements, and locomotor activity studies because data collection for these studies is automated and confers high objectivity.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☐	☒	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	Nissl: ThermoFisher Scientific, N21479 UCP1: Abcam, ab10983 Alpha-Tubulin: Abcam, ab4074 Tyrosine Hydroxylase (TH): Abcam, ab113 Hoechst 33342: ThermoFisher Scientific, 62249 Isolectin B4: ThermoFisher Scientific, I21411 Insulin: Dako, A0564
Validation	All antibodies are commercial in origin. Validation statements can be found on the manufacturer's website for the following: Fluorescent Nissl (ThermoFisher Scientific, N21479): https://www.thermofisher.com/order/catalog/product/N21479#/N21479 UCP1 (Abcam, ab10983): https://www.abcam.com/ucp1-antibody-ab10983.html Alpha-Tubulin (Abcam, ab4074): https://www.abcam.com/alpha-tubulin-antibody-loading-control-ab4074.html Tyrosine Hydroxylase (Abcam, ab113): https://www.abcam.com/tyrosine-hydroxylase-antibody-ab113.html Hoechst 33342 (ThermoFisher Scientific, 62249): https://www.thermofisher.com/order/catalog/product/62249#/62249 Isolectin B4 (ThermoFisher Scientific, I21411): https://www.thermofisher.com/order/catalog/product/I21411?SID=srch-hj-I21411#/

I21411?SID=srch-hj-I21411
Insulin (Dako, A0564): https://www.agilent.com/en/product/immunohistochemistry/antibodies-controls/primary-antibodies/insulin-(autostainer-link-48)-76277

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Species: Mus musculus, Opn5 KO, Opn5-cre, Rx-cre, and Lepr-cre; Opn5fl/fl cKO were on a mixed (CD1/FVB/129/C57) background. Full spectrum and minus violet mice were on a C57BL/6J background. ROSA26::FLPeR breeder mice were on a 129S4/Sv background. E2a-cre breeder mice were on a mixed C57BL/6J and C57BL/6N background. Both male and female mice at ages of postnatal day (P)8, P16, P21, P35, P60, P70, and P90 - P120 were used in the study.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve field-collected samples.
Ethics oversight	Cincinnati Children's Hospital Medical Center IACUC

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Adaptable haemodynamic endothelial cells for organogenesis and tumorigenesis

https://doi.org/10.1038/s41586-020-2712-z

Received: 6 December 2017

Accepted: 8 June 2020

Brisa Palikuqi[1], Duc-Huy T. Nguyen[1], Ge Li[1], Ryan Schreiner[1,2], Alessandro F. Pellegata[3], Ying Liu[1], David Redmond[1], Fuqiang Geng[1], Yang Lin[1], Jesus M. Gómez-Salinero[1], Masataka Yokoyama[1], Paul Zumbo[4], Tuo Zhang[5], Balvir Kunar[1], Mavee Witherspoon[6], Teng Han[6], Alfonso M. Tedeschi[3], Federico Scottoni[3], Steven M. Lipkin[6], Lukas Dow[6], Olivier Elemento[7], Jenny Z. Xiang[5], Koji Shido[1], Jason R. Spence[8], Qiao J. Zhou[1], Robert E. Schwartz[1,9], Paolo De Coppi[3,10], Sina Y. Rabbany[1,11] & Shahin Rafii[1]✉

Endothelial cells adopt tissue-specific characteristics to instruct organ development and regeneration[1,2]. This adaptability is lost in cultured adult endothelial cells, which do not vascularize tissues in an organotypic manner. Here, we show that transient reactivation of the embryonic-restricted ETS variant transcription factor 2 (ETV2)[3] in mature human endothelial cells cultured in a serum-free three-dimensional matrix composed of a mixture of laminin, entactin and type-IV collagen (LEC matrix) 'resets' these endothelial cells to adaptable, vasculogenic cells, which form perfusable and plastic vascular plexi. Through chromatin remodelling, ETV2 induces tubulogenic pathways, including the activation of RAP1, which promotes the formation of durable lumens[4,5]. In three-dimensional matrices—which do not have the constraints of bioprinted scaffolds—the 'reset' vascular endothelial cells (R-VECs) self-assemble into stable, multilayered and branching vascular networks within scalable microfluidic chambers, which are capable of transporting human blood. In vivo, R-VECs implanted subcutaneously in mice self-organize into durable pericyte-coated vessels that functionally anastomose to the host circulation and exhibit long-lasting patterning, with no evidence of malformations or angiomas. R-VECs directly interact with cells within three-dimensional co-cultured organoids, removing the need for the restrictive synthetic semipermeable membranes that are required for organ-on-chip systems, therefore providing a physiological platform for vascularization, which we call 'Organ-On-VascularNet'. R-VECs enable perfusion of glucose-responsive insulin-secreting human pancreatic islets, vascularize decellularized rat intestines and arborize healthy or cancerous human colon organoids. Using single-cell RNA sequencing and epigenetic profiling, we demonstrate that R-VECs establish an adaptive vascular niche that differentially adjusts and conforms to organoids and tumoroids in a tissue-specific manner. Our Organ-On-VascularNet model will permit metabolic, immunological and physiochemical studies and screens to decipher the crosstalk between organotypic endothelial cells and parenchymal cells for identification of determinants of endothelial cell heterogeneity, and could lead to advances in therapeutic organ repair and tumour targeting.

Endothelial cells (ECs) in zonated capillaries sustain tissue-specific homeostasis and supply angiocrine factors to guide organ regeneration[1,2]. By contrast, maladaptation of ECs contributes to fibrosis and tumour progression[6,7]. The mechanism(s) by which ECs acquire adaptive tissue-specific heterogeneity or maladapt within the scarred tissues or tumour microenvironment are unknown. Identifying the molecular determinants of vascular heterogeneity requires the generation of malleable and perfusable vascular networks that are

---

[1]Division of Regenerative Medicine, Ansary Stem Cell Institute, Department of Medicine, Weill Cornell Medicine, New York, NY, USA. [2]Department of Ophthalmology, Margaret Dyson Vision Research Institute, Weill Cornell Medicine, New York, NY, USA. [3]Stem Cell and Regenerative Medicine Section, DBC Programme, Great Ormond Street Institute of Child Health, University College London, London, UK. [4]Applied Bioinformatics Core, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA. [5]Genomics Resources Core Facility, Weill Cornell Medicine, New York, NY, USA. [6]Sandra and Edward Meyer Cancer Center, Weill Cornell Graduate School of Medical Sciences, Departments of Biochemistry and Medicine, Weill Cornell Medicine, New York, NY, USA. [7]Caryl and Israel Englander Institute for Precision Medicine, Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA. [8]Department of Internal Medicine, University of Michigan School of Medicine, Ann Arbor, MI, USA. [9]Department of Physiology, Biophysics and Systems Biology, Weill Cornell Medicine, New York, NY, USA. [10]Specialist Neonatal and Paediatric Surgery, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK. [11]Bioengineering Program, DeMatteis School of Engineering and Applied Science, Hofstra University, Hempstead, NY, USA. ✉e-mail: srafii@med.cornell.edu

responsive and can conform to microenvironmental and biophysical signals[8].

Attempts to uncover the crosstalk between adult ECs and non-vascular cells—for example, through the generation of decellularized scaffolds[9,10], organ-on-chip models[11,12] and three-dimensional (3D) bioprinting, as well as the culturing of normal[13] and malignant organoids[14]—have met with hurdles. In these approaches, ECs do not have the cellular freedom to directly interact with parenchymal and tumour cells, owing to the physical constraints that are imposed by artificial semipermeable biomaterials used in organ-on-chip systems and low-volume microfluidic devices, and the lack of adaptive ECs[11]. Moreover, the use of non-physiological matrices such as Matrigel, poses challenges for translation to the clinic. Thus, transcriptional resetting of adult human ECs to generate adaptable tubulogenic and perfusable ECs in defined matrices will provide insights into vascular diversity and therapeutic organ regeneration.

During development, ETV2 functions as a pioneer transcription factor that induces vascular cell fate and lumen morphogenesis[3,15]. ETV2 is expressed in ECs during vasculogenesis, but is turned off mid-gestation, when the primitive capillary networks are established[15], and is not expressed in adult ECs. Transient reintroduction of ETV2 into parenchymal cells induces a stable EC fate[16]. Here, we show that in addition to specifying vascular cell fate, transient reactivation of ETV2 resets mature adult human vascular ECs (VECs) to embryonic-like malleable vasculogenic ECs, hereafter referred to as 'reset VECs' (R-VECs). R-VECs self-organize into adaptable, large-volume 3D lumenized vascular networks that can transport human blood and physiologically arborize decellularized tissues, islets and normal and malignant organoids, and that can build durable capillaries in vivo.

R-VECs form stable vessels in vitro

Human ECs transduced with lentivirus to express ETV2, in serum-free medium form functional, durable and adaptable 3D vessels by transitioning through three stages (Fig. 1a, Extended Data Fig. 1a). In the first (induction) stage, ETV2 upregulates vasculogenic and tubulogenic factors in flat EC cultures. During the second (remodelling) stage, R-VECs that are placed in matrices, self-assemble into sprouting and lumenized 3D vessels. At the third stage, R-VECs are non-proliferative, maintaining stabilized and adaptive 3D patterned capillaries (Extended Data Fig. 1b).

Human umbilical vein ECs (HUVECs) transduced with ETV2 showed a 50-fold increase in the area of vessels formed over 8 weeks compared to naive HUVECs, which did not form durable vessels in any of the enriched angiogenic media that we tested (Fig. 1b, c, Extended Data Fig. 1c, d, Supplementary Video 1a). In addition, mature adult human EC populations isolated from adipose, cardiac, aortic and dermal tissues and transduced with ETV2 formed long-lasting and patterned R-VEC plexi (Extended Data Fig. 1e–g). Next, we investigated whether R-VEC vessel formation could be achieved without Matrigel. We identified a stoichiometrically defined ratio of laminin, entactin and type-IV collagen that is sufficient for the self-assembly of R-VECs into lumenized vessels similar to those formed in Matrigel (Fig. 1d, Extended Data Fig. 1h). This composite matrix of laminin, entactin and type-IV collagen is hereafter referred to as LEC matrix. Confocal and electron microscopy showed that R-VECs organized into vessels that exhibit a continuous, patent (open) lumen with the apicobasal polarity on both Matrigel and LEC matrix (Fig. 1e, Extended Data Fig. 1i). Moreover, transduction of ETV2 reduces stiffness in adult ECs (as measured by atomic force microscopy; AFM), which facilitates lumen formation (Extended Data Fig. 1j). To assess whether this tubulogenesis-promoting activity is specific to ETV2, we transduced HUVECs with a lentiviral vector expressing another ETS transcription factor, ETS1; and to test whether the survival of ECs could drive tubulogenesis, HUVECs were also transduced with a lentiviral construct of constitutively active myristoylated AKT1 (myrAKT1). Neither ETS1 nor myrAKT1 reset ECs to form stable vessels (Extended Data Fig. 1k–m).

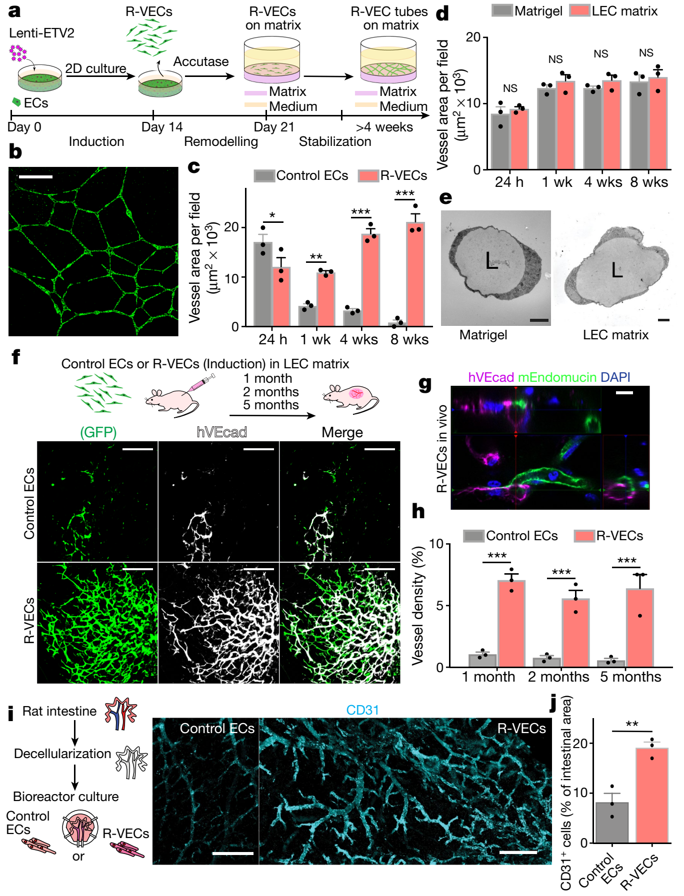


Fig. 1 | R-VECs self-assemble into 3D durable vessels in vitro and in vivo. **a**, Experimental set-up for vessel formation. A total of 10^5 control ECs or R-VECs were plated on Matrigel in serum-free StemSpan tube-formation medium (Supplementary Data 2). Lenti-EVT2, lentiviral ETV2 expression construct. **b**, *Z*-stack of GFP$^+$ R-VEC vessels at week 16. Scale bar, 1,000 µm. **c**, Quantification of tube formation in control ECs (HUVECs) and R-VECs (HUVEC-ETV2). **d**, Quantification of R-VEC vessels on Matrigel or LEC matrix. **e**, Electron microscopy images of stage-3 vessels on Matrigel and LEC matrix. L, lumen. Scale bars, 5 µm. **f**, Top, schematic of in vivo plug experiment in which control ECs or R-VECs fluorescently labelled with GFP were subcutaneously injected as a single-cell LEC suspension into SCID-beige mice. Bottom, whole-mount confocal images of R-VEC plugs and control EC plugs at five months. A fluorescently labelled antibody against human VEcadherin (hVEcad) was injected retro-orbitally before mice were euthanized. Scale bars, 200 µm. **g**, Orthogonal projection showing the anastomosis of mouse vessels and human VEcad$^+$ vessels. Sections were post-stained for mouse endomucin (mEndomucin). Scale bar, 10 µm. **h**, Quantification of the density of human vessels in the plugs, defined as the percentage of GFP positive vessels of the scanned area. **i**, **j**, Experimental procedure for the decellularized intestine cultures (**i**, left). R-VECs repopulated the vasculature, lining blood vessels including the distal capillaries. At day 7 the bioreactors were stained for human CD31, imaged (**i**, right) and quantified (**j**). Scale bars, 500 µm. Data are mean ± s.e.m. NS, not significant; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. For statistics, see Supplementary Data 1.

We quantified the mRNA and protein levels of ETV2 in R-VECs from stages 1 to 3 (Extended Data Fig. 2a–d). ETV2 protein levels peaked during stage 2 but were downregulated by more than 90% at stage 3, which could not be accounted for by the minor drop in *ETV2* mRNA levels (Extended Data Fig. 2a–d). Treatment with the proteasome inhibitor MG132 at stage 3 restored ETV2 protein levels by sixfold—approaching its original expression levels—which indicates that proteasomal proteolysis regulates ETV2 expression (Extended Data Fig. 2e, f).

To examine whether short-term induction of ETV2 is sufficient to generate R-VECs, we used a reverse tetracycline-controlled transactivator (rtTA) doxycycline-inducible system, in which doxycycline induces the expression of ETV2 (induced R-VECs; iR-VECs) (Extended Data Fig. 2g, h). Induction of ETV2 was transiently required until the first week of stage 2; after that, iR-VEC vessels sustain their stability without continuous ETV2 induction (Extended Data Fig. 2i–k).

Thus, short-term expression of ETV2 confers adult ECs with the capacity to self-assemble into stable and durable patterned vessels, without affecting cell survival and proliferation and without the physical constraints of artificial bioprinted scaffolds and restrictive synthetic barriers.

R-VECs form durable vessels in vivo

SCID-beige mice were implanted subcutaneously with mCherry- or GFP-labelled control human ECs or R-VECs suspended in LEC matrix. One to five-months after implantation, R-VECs—but not control ECs—self-organized into long-lasting, branching and patterned vessels in vivo. Injection of R-VEC-implanted mice with an antibody directed against human vascular endothelial cadherin (VEcad) showed that R-VEC vessels anastomose to the endomucin-positive mouse vasculature, establishing a mosaic of functional perfused vessels throughout the plug (Fig. 1f–h, Extended Data Fig. 3a). Mouse perivascular cells wrap around R-VEC vessels, with larger arterioles covered with a thicker layer of smooth muscle cells and less coverage in smaller capillaries (Extended Data Fig. 3b, c). iR-VECs also assembled into stable vessels in LEC matrix, and one week of doxycycline treatment in vivo was sufficient to retain vascular stability (Extended Data Fig. 3d, e). The lack of extravasation of intravenously injected 70-kDa dextran in mice indicated that R-VEC and iR-VEC vessels in in vivo plugs were non-leaky and patent. By contrast, human ECs that were transduced with KRAS formed leaky and disorganized vessels, reminiscent to those of haemangiomas (Extended Data Fig. 3f). Unlike implants of KRAS-transduced endothelial cells, R-VEC implants did not exhibit aberrant growth, haemangiomas or tumours, and they retained perfused and organized vessels for 10 months (Extended Data Fig. 4a–e). In summary, R-VECs build durable, anastomosed and pericyte-covered capillaries that are structurally normal and show no signs of vascular anomalies or tumours.

R-VECs arborize decellularized scaffolds

We next examined whether R-VECs can functionally populate the denuded vascular lining of decellularized tissues. Although large vessels in decellularized scaffolds can be colonized with ECs, it is challenging to vascularize the abundant smaller capillaries[9]. Stage-1 R-VECs, but not control ECs, fully populated the narrow small capillaries evenly throughout the decellularized rat intestine scaffolds ex vivo (Fig. 1i, j, Extended Data Fig. 5a–d). After one week of ex vivo culture, the revascularized intestinal explants were implanted in the omentum of immunocompromised mice. Intravital anti-human VEcad staining at one and four weeks showed that R-VEC-vascularized scaffolds retained their patency and anastomosed to the mouse vasculature (Extended Data Fig. 5e). At four weeks, R-VEC vessels persisted at a higher rate in vivo compared to naive control ECs, owing to their integrity and low rate of apoptosis (Extended Data Fig. 5f, g). Thus, R-VECs enable the functional arborization of decellularized tissues for therapeutic regeneration.

ETV2 remodels ECs to primitive plexi

To uncover the mechanism by which ETV2 drives vascular resetting, we performed RNA sequencing (RNA-seq) in stage-1 R-VECs and control ECs (Fig. 2a–c). Gene Ontology (GO) analyses revealed the upregulation of genes in pathways that regulate vasculogenesis, angiogenesis, GTPase activity, extracellular matrix remodelling and the response to mechanical stimuli (Fig. 2b, c, Extended Data Fig. 6a, b). At stage 1, R-VECs maintain their vascular identity by sustaining the expression of EC-specific genes (Extended Data Fig. 6c). After ETV2 induction, a group of 490 genes was differentially expressed among various tissue-specific

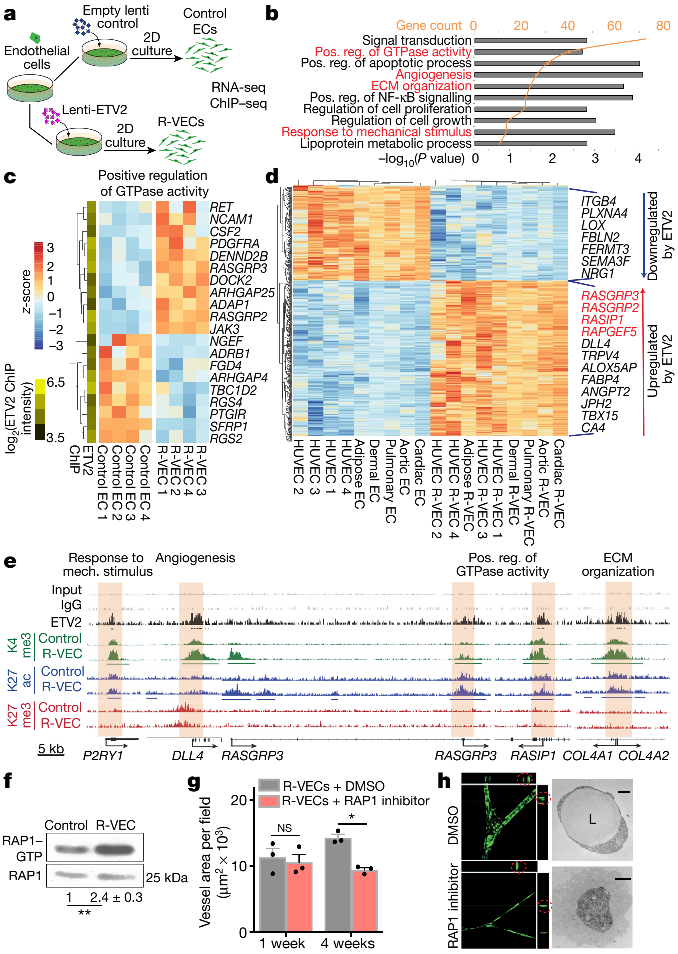


Fig. 2 | Transcriptome and epigenetic analyses of R-VEC signatures. **a**, Schematic of RNA-seq and ChIP–seq performed in the induction phase (day 14) on R-VECs and control ECs. **b**, RNA-seq of R-VECs or control HUVECs in stage 1 (2D monolayers). GO term analysis was performed on differentially expressed genes. GO categories are ordered on the basis of the number of differentially expressed genes. Heat maps for GO categories in red are presented in Fig. 2c and Extended Fig. 6b. ECM, extracellular matrix; Pos. reg., positive regulation. **c**, Heat map of genes in one top GO category. Values are $\log_2$-normalized counts per million (CPM), centred and scaled by row. ETV2 binding from ChIP–seq at the promoter of each differentially expressed gene is shown in the yellow-and-green heat map (left). **d**, Heat map of 490 differentially expressed genes across ECs of different tissues (stage 1, induction phase) upon ETV2 expression. Tissue-adjusted $\log_2$-transformed CPM, centred and scaled by row. **e**, ETV2 ChIP–seq in R-VECs during the induction phase (stage 1; 2D) using an anti-Flag antibody or mouse IgG as control. ChIP for H3K4me3, H3K27ac and H3K27me3 was performed in both control ECs and R-VECs at stage 1. Enriched regions were analysed by ChIP–seq. Horizontal bars underneath peaks represent significantly changed regions. Promoter regions bound by ETV2 are highlighted in cream. Track range ETV2/K27me3/K27ac, 0–0.3; K4me3/input/IgG, 0–1. **f**, Western blot for active RAP1–GTP compared to total RAP1 input for stage 1 2D control ECs (HUVECs) and R-VECs (HUVEC-ETV2). The quantification of RAP1–GTP compared to total RAP1 is shown below the blot and presented as mean ± s.e.m. **g**, Quantification of R-VEC vessel formation after treatment with RAP1 inhibitor or dimethyl sulfoxide (DMSO). **h**, *Z*-stack confocal images and electron microscopy images of R-VEC vessels treated with RAP1 inhibitor or DMSO at four weeks. Red circles indicate orthogonal cross-sections. Scale bars, 5 μm (top); 2 μm (bottom). Data are mean ± s.e.m. NS, not significant; $^{*}P < 0.05$, $^{**}P < 0.01$. For statistics, see Supplementary Data 1.

adult human ECs, including cardiac, dermal, aortic, pulmonary and adipose-derived R-VECs (Fig. 2d, Extended Data Fig. 6d). Chromatin immunoprecipitation followed by sequencing (ChIP–seq) analysis of K4me3, K27ac and K27me3 histone modifications in both R-VECs and control ECs showed that ETV2 bound to the promoters of several differentially expressed vascular-specific genes—and to the promoters of pro-tubulogenesis genes, which are silenced in mature ECs (Fig. 2c, e, Extended Data Fig. 6e–h). Therefore, ETV2 resets the chromatin and transcriptome of mature ECs through the direct reactivation of suppressed tubulogenic and vasculogenic genes.

After ETV2 transduction, genes encoding Ras-interacting protein 1 (*RASIP1*) and three guanine nucleotide exchange factors (GEFs) that are involved in the activation of the small GTPase RAP1 (*RASGRP2*, *RASGRP3* and *RAPGEF5*)—all of which are crucial for lumen formation[4,5]—were upregulated in all tissue-specific ECs (Fig. 2c, d). Similarly, differential expression of genes in the RAP1 pathway was found in ETV2-positive ECs isolated from ETV2–Venus reporter mouse embryos at embryonic stage 9.5 (E9.5) (Extended Data Fig. 7a, b). ChIP–seq analysis of stage-1 R-VECs confirmed the direct binding of ETV2 to *RASGRP3* and *RASIP1* promoters and a subsequent increase in K4me3 and K27ac histone marks at these genes (Fig. 2e). A pull-down of active RAP1–GTP in stage-1 R-VECs showed that the levels of active RAP1–GTP were higher in R-VECs than in naive ECs (Fig. 2f). Vessel formation was reduced and no lumen was present after treatment with the RAP1 inhibitor GGTI-298 (Fig. 2g, h). Similarly, knockdown of *RASGRP3* by short hairpin RNA (shRNA) disrupted R-VEC-mediated tubulogenesis (Extended Data Fig. 7c). Therefore, ETV2 potentiates lumen formation in part through the upregulation of RAP1 GEFs.

In vitro, stage-3 R-VECs upregulate the expression of genes that are involved in mechanosensing (*PIEZO2*, *KLF2* and *KLF4*) and EC remodelling (*ATF3*), which are not expressed in cultured mature ECs (Extended Data Fig. 7d). We confirmed this result by isolating R-VECs from in vivo plugs and comparing their transcriptome to that of freshly isolated HUVECs and stage-3 R-VEC stable vessels (Extended Data Fig. 7d). Notably, the genes upregulated in stage-3 R-VECs (*PIEZO2*, *KLF2* and *KLF4*) were bound by ETV2 and epigenetically primed for expression in stage-1 two-dimensional (2D) R-VECs (Extended Data Fig. 7e). Thus, ETV2 resets the chromatin landscape of mature ECs to an in vivo physiological configuration that is responsive and conforms to microenvironmental cues—reminiscent of generic vasculogenic ECs.

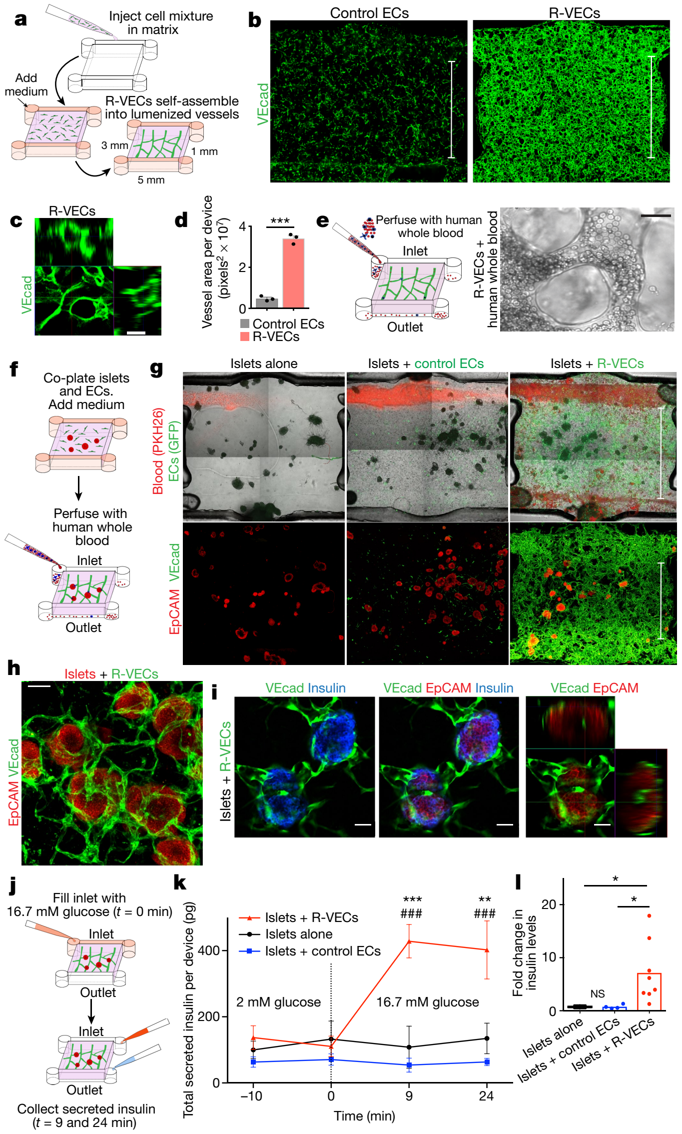


Fig. 3 | R-VECs haemodynamically and physiologically vascularize human islets. a, Overview of microfluidic device measuring 5 × 3 × 1 mm and holding 15 µl fibrin gel. **b**, Representative images of devices with control ECs or R-VECs stained with human VEcad antibody at day 7. Scale bars, 3 mm. **c**, Orthogonal representation of intact lumen formation in R-VECs. Scale bar, 50 µm. **d**, Quantification of vessel area in devices with control ECs versus R-VECs. **e**, Intact heparinized human peripheral blood (100 µl) composed of a full complement of red blood cells, white blood cells, platelets and unperturbed plasma was injected and perfused through the R-VEC vessels. Right, representative live image of blood flow through R-VECs (see also Supplementary Video 1b, c). Scale bar, 25 µm. **f**, Experimental set-up for co-seeding human islets with control ECs or R-VECs in microfluidic devices. **g**, Fluorescently labelled human heparinized whole blood (red, PKH26 red fluorescent dye) was perfused through the microfluidic devices (day 4) (see also Supplementary Video 2b–d). *Z*-stack projections of whole devices of islet explants post-stained with EpCAM and VEcad (day 4). Scale bars, 3 mm. **h**, Magnified area of direct interaction of R-VECs with co-cultured islets in a microfluidic device. Scale bar, 100 µm. **i**, Single section and orthogonal projection of human islets vascularized by R-VECs in a microfluidic device. Scale bars, 50 µm. **j**, Experimental set-up for the glucose-stimulation test in microfluidic devices. **k**, Insulin levels were measured at 2 mM glucose ($t = -10$ and 0 min, basal level) and 9 and 24 min after stimulation with 16.7 mM glucose. * represents statistical tests versus islets alone; # represents statistical tests versus islets + control ECs. **l**, Fold change in insulin levels at the outlet (insulin levels at 16.7 mM/insulin levels at 2 mM), 9 min after high-glucose stimulation. Data are mean ± s.e.m. NS, not significant; $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{\#\#\#}P < 0.001$. For statistics, see Supplementary Data 1.

R-VECs build haemodynamic vessels

We tested the capacity of R-VEC vessels to self-congregate into sprouting vascular networks in the absence of pre-patterned scaffolds and synthetic barriers and to sustain a laminar flow in vitro in large-volume microfluidic devices. R-VECs or control ECs were seeded in a 5 × 3 × 1-mm microfluidic device that can accommodate more than 45,000 stage-1 ECs within a 15-µl volume of fibrin gel[17] (Fig. 3a). Within three days, R-VECs self-organized into a multilayered, branching and interconnected vascular plexus, maintaining their 3D lumenized stability (Fig. 3b–d). Notably, R-VEC vessels allowed the gravity-driven transport of heparinized human whole peripheral blood with a full complement of plasma, platelets, white and red blood cells (Fig. 3e, Supplementary Video 1b, c). During the transport of blood, R-VEC capillary networks sustained their vascular integrity and were haemodynamically stable from the inlet to the outlet chambers of the microfluidic device, enduring the force of blood flow without collapse, regression or thrombosis. R-VECs therefore maintain the haemodynamic vascularization of tissues, and pave the way for an Organ-On-VascularNet platform (Supplementary Video 1d).

R-VECs physiologically vascularize islets

We assessed the potential of R-VECs to functionally vascularize human islets in the perfusable microfluidic devices. Currently, organ-on-chip

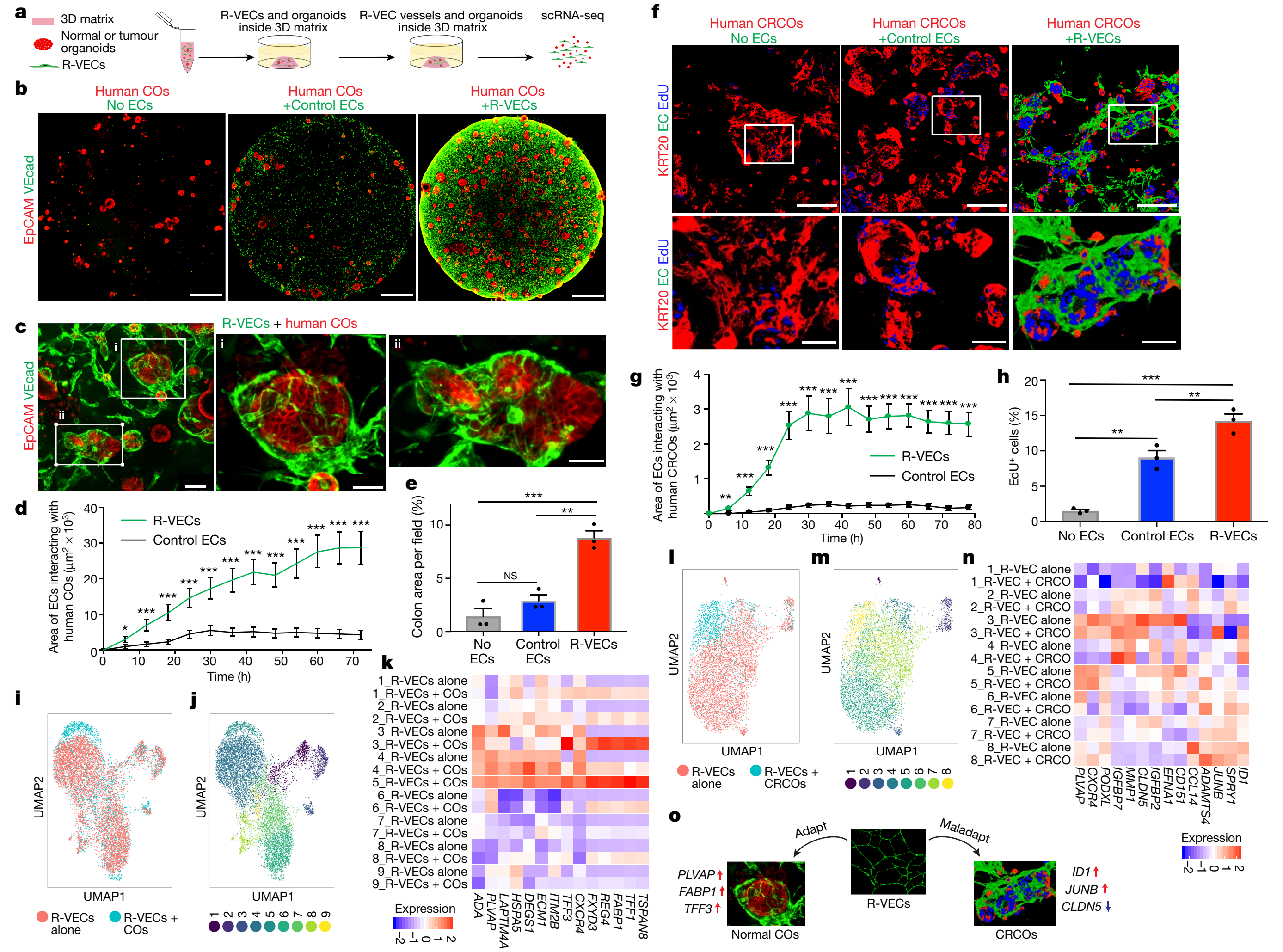


Fig. 4 | R-VECs arborize and conform to normal and tumour organoids. **a**, Control ECs or R-VECs were seeded with human normal colon organoids (COs) or colorectal cancer organoids (CRCOs) in Matrigel droplets. **b**, **c**, Confocal *Z*-projections (**b**) and magnified images (**c**) of human COs alone or co-cultured with control ECs or R-VECs on day 8. Scale bars, 1 mm (**b**); 100 µm (**c**, left image); 50 µm (**c**, right images). **d**, Surface area of control ECs or R-VECs interacting with human COs in LEC matrix, quantified over a 72-h time lapse. **e**, Quantification of colon area (as stained by EpCAM) per field in COs alone or co-cultured with control ECs or R-VECs on day 8. **f**, Human CRCOs were seeded with control ECs or R-VECs in Matrigel droplets. Confocal images of CRCOs alone or co-cultured with control ECs or R-VECs, post-stained for KRT20 at day 8 after a 4.5-h EdU pulse. Scale bars, 100 µm. **g**, Surface area of control ECs or R-VECs interacting with human CRCOs in Matrigel, quantified over a 78-h time lapse. **h**, Quantification of EdU$^+$ CRCO cells (in **f**) in CRCOs alone or CRCOs co-cultured with control ECs or R-VECs on day 8. **i**, **j**, R-VECs in single-cell suspension were cultured either alone or mixed with human COs and subjected to scRNA-seq on day 7. **i**, Uniform manifold approximation and projection (UMAP) plot for the EC fractions of R-VECs alone and R-VECs co-cultured with COs. **j**, UMAP plot of 9 unique clusters, in ECs from both the R-VECs alone group and the R-VECs +COs group. **k**, Heat map of differentially expressed genes from cluster 5, enriched among R-VECs but absent in R-VEC-alone cultures. **l**, **m**, R-VECs in single-cell suspension were cultured either alone or mixed with human CRCOs and subjected to scRNA sequencing on day 7. **l**, UMAP plot for the EC fractions of R-VECs alone and R-VECs co-cultured with CRCOs. **m**, UMAP plot of 8 unique clusters, in ECs from both the R-VECs alone group and the R-VECs + CRCOs group. **n**, Heat map of differentially expressed genes from cluster 8, enriched among R-VECs in culture with CRCOs. **o**, Schematic of the adaptation and maladaptation of R-VECs. Data are mean ± s.e.m. NS, not significant; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. For statistics, see Supplementary Data 1.

devices[11,12] segregate ECs from parenchymal cells with physical barriers, and are thus unsuitable for studying islets, which require active interaction with ECs to maintain their function[18]. We seeded around 40 human islets, alone or in the presence of control ECs or R-VECs, in 15-µl microfluidic devices (Fig. 3f). Within three days, R-VECs—but not control ECs—arborized islets with continuous 3D vascular networks, which extended deep into islets and metabolically irrigated insulin-secreting β-cells (Fig. 3g–i, Supplementary Video 2a–f). Heparinized human blood travelled through the R-VEC-co-opted islets, with intact haematopoietic cells perfusing the vascularized islets (Fig. 3g, Supplementary Video 2b–e).

We used a glucose-stimulation test to assess islet function (Fig. 3j), and found that islets arborized with R-VECs responded to high glucose by secreting insulin, as measured at the device outlet at 9 and 24 minutes of stimulation (Fig. 3k). There was a sevenfold increase in insulin secretion in glucose-stimulated R-VEC-co-opted islets, but not in control ECs or islet-alone cultures (Fig. 3l). Similar results were observed in co-cultured islet explants arborized by R-VECs in static Matrigel droplets (Extended Data Fig. 8a–e). Thus, R-VECs self-congregate in large-volume microfluidic devices into haemodynamically stable vessels that physiologically perfuse and sustain glucose-sensing human β-cells.

R-VECs vascularize organoids and tumoroids

We next assessed the capacity of R-VECs to functionally arborize organoids composed of healthy or malignant human cells, in order to model tissue- and tumour-specific adaptive responses of ECs and set the stage for organ regeneration. Normal colon organoids (COs) were established and maintained from healthy human colon crypts[19,20] (Extended Data Fig. 8f). Next, the COs were mixed with either control ECs or stage-1 R-VECs in static 50-µl droplets of Matrigel or LEC matrix (Fig. 4a). R-VECs sustained the arborization of the COs throughout the matrix droplet, with a higher vessel area than control ECs (Fig. 4b, c, Supplementary Video 3a). Moreover, as tracked in a 72-h time-lapse video, R-VECs interacted and engaged significantly more with the cells within the COs, as compared to control ECs (Fig. 4d, Supplementary Video 3b). The surface area of COs was larger in the presence of R-VECs, with no change in the differentiation of COs as assessed by the expression of stem and progenitor cell markers (Fig. 4e, Extended Data Fig. 8g). R-VECs also arborized mouse small intestinal organoids, with an increase in the vessel area and the number of R-VEC sprouts per organoid (Extended Data Fig. 8h–j). Thus, R-VECs instructively sustain the proliferation and integrity of COs, while preserving their differentiation status.

Tumour vasculature is composed of abnormal capillaries that supply aberrant factors that instigate tumour growth[7]. To determine whether R-VECs can acquire and report on the maladapted features of tumour vessels, we mixed stage-1 R-VECs with patient-derived colorectal cancer organoids (CRCOs) (Fig. 4f, Supplementary Video 3c). Within 24 hours, R-VECs, but not control ECs, migrated to and erratically infiltrated tumour organoids (Supplementary Video 3c). Similar to human COs, the vessel area in CRCOs mixed with R-VECs, and the interaction of R-VECs with CRCOs—as tracked in a 72-h time-lapse video—were increased compared to control ECs (Fig. 4f, g, Extended Data Fig. 8k, Supplementary Video 3c). Staining for the epithelial marker EpCAM revealed intimate cell–cell interactions between the tumoroids and the R-VECs, with a higher percentage of EdU-positive proliferating tumour cells in the R-VEC than the control EC co-cultures (Fig. 4f, h, Extended Data Fig. 8l). Hence, R-VECs establish an adaptive 3D vascular niche that can be used to decipher the crosstalk between ECs and normal or tumour organoids.

R-VECs adapt to organoids and tumoroids

We performed single-cell RNA-seq (scRNA-seq) on the 3D R-VEC-vascularized human COs or CRCOs to assess the adaptability of R-VECs. R-VECs were cultured alone or co-cultured with human COs or CRCOs for seven days, isolated and subjected to scRNA-seq using the 10X Genomics Chromium platform (Extended Data Figs. 9a, 10a). ECs were identified by their expression of *VEcad* (also known as *CDH5*) *CD31* (*PECAM1*) and *VEGFR2* (*KDR*) and epithelial cells by their expression of *EPCAM*, *CDH1* and *KRT19* (Extended Data Figs. 9b–e, 10b–e). The identity of the COs was validated by the expression of *SATB2*, *CA4* and *CA2*, among other genes (Extended Data Fig. 9f).

R-VECs that were co-cultured with malignant or normal organoids showed changes in their clustering patterns and gene expression when compared to R-VECs that were cultured alone (Fig. 4i–n). R-VECs that interacted with COs were enriched in EC organotypic marker genes, including *PLVAP* and *TFF3* (cluster 5, absent in R-VECs alone)[1,21] (Fig. 4i–k). By contrast, R-VECs that arborized CRCOs were enriched in clusters of genes with typical attributes of tumour ECs, including *ID1*, *JUNB* and *ADAMTS4* (cluster 8), whereas genes responsible for junctional integrity—such as *CLDN5* (cluster 5, cluster 7)—were selected against[22] (Fig. 4l–n). In response to association with R-VECs, colon tumour cells upregulated their expression of marker genes that are linked to poor prognosis and high rates of metastasis, including higher levels of *MSLN*[23], and downregulated their expression of *MT1G*, *MT1X* and *MT2A*[24] (Extended Data Fig. 10f–h). These data provide further evidence that R-VECs model an adaptable 3D vascular niche that responds to microenvironmental stimuli (Fig. 4o).

Discussion

We have created haemodynamic, self-organizing, large-volume 3D R-VEC vascular plexi in a Matrigel-free LEC matrix, which mimic primitive pliable blood vessels. R-VECs sustain their tubulogenic potential in diverse serum-free media compositions enabling the functional vascularization of organoids and tissue explants, notably islets. These networks do not have the constraints of synthetic scaffolds and semipermeable membranes, and allow the direct cellular interaction of ECs with parenchymal and tumour cells. Transient reintroduction of ETV2—which is silenced during fetal development—into adult human ECs induces a molecular reset of cell tubulogenic and adaptability attributes that are lost in cultured mature ECs[4,5]. In R-VECs, the RAP1 pathway is activated through RAP1 GEFs and the RASIP1 effector, allowing lumen formation in a flow- and pericyte-independent manner. ETV2 resets the vasculogenic memory to an early embryonic stage, and thereby renders R-VECs receptive to microenvironmental cues[1,2]. In stabilized R-VEC vessels, the expression of ETV2 was spontaneously reduced through proteasomal proteolysis, suggesting that transient expression of ETV2 is sufficient to reset ECs into a plastic and adaptive state.

The capacity of R-VECs to self-assemble into perfusable vascular networks that can transport human blood enables the 3D physiological vascularization of scalable and organ-level micro- and macrofluidic devices. This licenses R-VECs to recapitulate the physiochemical and multicellular geometry of blood-perfusable vascular niches that, by deploying angiocrine factors, directly enhance the frequency of co-cultured organoids. In addition, R-VECs conform to signals that are produced by organoids or tumoroids and, reciprocally, tumour cells upregulate markers that are associated with poor outcomes in response to signals induced by subverted R-VECs. Our R-VEC Organ-On-VascularNet platform therefore overcomes the constraints of costly, technically challenging and non-physiological organ-on-chip models, the design of which prevents the direct cellular interaction of ECs with non-vascular cells.

Co-cultures of organoids with blood-perfusable pericyte-coated R-VECs could serve as a tissue-specific biological platform for the delivery of engineered immune cells (such as CAR-T cells) and chemotherapeutic agents, and could also be used to unravel the pathogenesis of microangiopathy in diseases such as coronavirus disease 2019 (COVID-19). The durable tubulogenic capacity, scalability, haemodynamic blood perfusibility, geometrical malleability, medium compatibility and cellular adaptability of R-VECs—which are capable of vascularizing normal and malignant organoids, as well as decellularized scaffolds—will lay the foundation for physiological, metabolic and immunological studies and pharmaceutical screening. The R-VEC Organ-On-VascularNet model permits the construction of functional and perfused implantable tissues ex vivo, opening a new chapter in translational vascular medicine for tissue-specific regeneration and for targeting the corrupted vascular niches of tumours.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2712-z.

1. Augustin, H. G. & Koh, G. Y. Organotypic vasculature: from descriptive heterogeneity to functional pathophysiology. *Science* **357**, eaal2379 (2017).
2. Rafii, S., Butler, J. M. & Ding, B. S. Angiocrine functions of organ-specific endothelial cells. *Nature* **529**, 316–325 (2016).
3. Lee, D. et al. ER71 acts downstream of BMP, Notch, and Wnt signaling in blood and vessel progenitor specification. *Cell Stem Cell* **2**, 497–507 (2008).
4. Barry, D. M. et al. Rasip1-mediated Rho GTPase signaling regulates blood vessel tubulogenesis via nonmuscle myosin II. *Circ. Res.* **119**, 810–826 (2016).

5. Strilić, B. et al. The molecular basis of vascular lumen formation in the developing mouse aorta. *Dev. Cell* **17**, 505–515 (2009).
6. Carmeliet, P. & Jain, R. K. Molecular mechanisms and clinical applications of angiogenesis. *Nature* **473**, 298–307 (2011).
7. Cao, Z. et al. Molecular checkpoint decisions made by subverted vascular niche transform indolent tumor cells into chemoresistant cancer stem cells. *Cancer Cell* **31**, 110–126 (2017).
8. Nolan, D. J. et al. Molecular signatures of tissue-specific microvascular endothelial cell heterogeneity in organ maintenance and regeneration. *Dev. Cell* **26**, 204–219 (2013).
9. Pellegata, A. F., Tedeschi, A. M. & De Coppi, P. Whole organ tissue vascularization: engineering the tree to develop the fruits. *Front. Bioeng. Biotechnol*. **6**, 56 (2018).
10. Giobbe, G. G. et al. Extracellular matrix hydrogel derived from decellularized tissues enables endodermal organoid culture. *Nat. Commun*. **10**, 5658 (2019).
11. Ronaldson-Bouchard, K. & Vunjak-Novakovic, G. Organs-on-a-chip: a fast track for engineered human tissues in drug development. *Cell Stem Cell* **22**, 310–324 (2018).
12. Bhatia, S. N. & Ingber, D. E. Microfluidic organs-on-chips. *Nat. Biotechnol*. **32**, 760–772 (2014).
13. Lancaster, M. A. & Knoblich, J. A. Organogenesis in a dish: modeling development and disease using organoid technologies. *Science* **345**, 1247125 (2014).
14. Tuveson, D. & Clevers, H. Cancer modeling meets human organoid technology. *Science* **364**, 952–955 (2019).
15. Koyano-Nakagawa, N. & Garry, D. J. Etv2 as an essential regulator of mesodermal lineage development. *Cardiovasc. Res*. **113**, 1294–1306 (2017).
16. Ginsberg, M. et al. Efficient direct reprogramming of mature amniotic cells into endothelial cells by ETS factors and TGFβ suppression. *Cell* **151**, 559–575 (2012).
17. Nguyen, D. H. et al. Biomimetic model to reconstitute angiogenic sprouting morphogenesis in vitro. *Proc. Natl Acad. Sci. USA* **110**, 6712–6717 (2013).
18. Eberhard, D., Kragl, M. & Lammert, E. 'Giving and taking': endothelial and beta-cells in the islets of Langerhans. *Trends Endocrinol. Metab*. **21**, 457–463 (2010).
19. Sato, T. et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).
20. Miyoshi, H. & Stappenbeck, T. S. In vitro expansion and genetic modification of gastrointestinal stem cells in spheroid culture. *Nat. Protocols* **8**, 2471–2482 (2013).
21. Stan, R. V. et al. The diaphragms of fenestrated endothelia: gatekeepers of vascular permeability and blood composition. *Dev. Cell* **23**, 1203–1218 (2012).
22. Lyden, D. et al. Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature* **401**, 670–677 (1999).
23. Li, S. et al. Plasma mesothelin as a novel diagnostic and prognostic biomarker in colorectal cancer. *J. Cancer* **8**, 1355–1361 (2017).
24. Si, M. & Lang, J. The roles of metallothioneins in carcinogenesis. *J. Hematol. Oncol*. **11**, 107 (2018).

Methods

Cell culture of ECs

Approval for the use of discarded left-over HUVECs and human adipose tissue ECs was obtained through the Weill Cornell Medicine Institutional Review Board (IRB). The ECs were isolated in the laboratory as previously described, using a collagenase-based digestion approach[25,26]. The cells were then grown in tissue culture dishes coated with 0.2% gelatin in complete EC medium. Complete EC medium is composed of 400 ml M199, 100 ml heat-inactivated fetal bovine serum (FBS), 7.5 ml HEPES, 5 ml antibiotics (Thermo Fisher Scientific, 15070063), 5 ml glutamax (Thermo Fisher Scientific, 35050061), 5 ml lipid mixture (Thermo Fisher Scientific, 11905031), and 25 mg EC growth supplement (Alpha Aesar, J64516-MF) (Supplementary Data 2). The cells were transduced with lenti-PGK-ETV2 or an empty lentiviral vector at passage 1–2. In some instances, the cells were also labelled by using PGK-mCherry or PGK-GFP lenti-viral vectors. The cells were split 1:2 using accutase and passaged on gelatinized plates. As required, cells in 2D (stage 1, induction) were frozen down to be used in future experiments. Comparisons for all assays and co-cultures were performed using the same parental EC line lentivirally transduced with and without ETV2. Overall, HUVECs from more than 10 different isolations were used for the experiments. R-VECs used for tube-formation assays were of passage 5–10.

Human adipose-derived ECs were isolated by mechanical fragmentation followed by collagenase digestion for 30 min. After plating the crude population of cells on the plastic dish and expansion for 5 to 7 days, the cells were then sorted to purify VEcad$^+$CD31$^+$ ECs and expanded as described above. Human adipose ECs were cultured in the same medium as that described above for HUVECs. At least three different isolations of adipose ECs were used in our experiments. Human microvascular cardiac (PromoCell, C12286), aortic (PromoCell, C12272), pulmonary (PromoCell, C-12282) and microvascular dermal (PromoCell, C12265) ECs were acquired from PromoCell and cultured in EC growth medium MV (PromoCell, C22020).

Lentiviral transduction of ECs

ECs were transduced with ETV2 lenti-particles or empty vector lenti-particles. *ETV2* cDNA (NM_014209.3) was introduced into the pCCL-PGK lentivirus vector (Genecopeia). For ChIP analysis, a triple Flag tag was subcloned in the ETV2 construct at the amino terminus[27]. After one week of transduction, ECs were collected for mRNA isolation and quantitative PCR with reverse transcription (qRT–PCR) analysis. The relative *ETV2* RNA unit was determined by calculating the relative *ETV2* mRNA expression compared to *GAPDH* using the following formula: $(2^{-[\mathrm{Ct(ETV2)}-\mathrm{Ct(GAPDH)}]})\times 1{,}000$. Primers are found in Supplementary Data 4). Cells with a relative *ETV2* RNA unit within the range of 60–100 were used for all experiments. A multiplicity of infection (MOI) of 3 gave us relative expression levels of 60–80 as calculated by mRNA expression. MOI was calculated by converting particles of antigen P24 to infectious units per ml (IFU) and then to MOI based on cell number (kit: Katara, 632200). An MOI of 3 was found to be adequate for cardiac and aortic ECs; an MOI of 6 was required for adipose and dermal ECs. Polybrene at 2 μg ml^{-1} was used for all transductions. *ETS1*, *myrAKT*, *mCherry* and *GFP* were also introduced into the pCCL-PGK lentivirus vector and an MOI of 3 was used for all transductions.

For inducible expression of ETV2, ECs were transduced with doxycycline-inducible ETV2 lenti-viruses (pLV[Exp]-Puro-TRE > hETV2 (NM_014209.3), VectorBuilder VB170514-1062dfs and pLV[Exp]-Neo-CMV > tTS/rtTA_M2, VectorBuilder VB160419-1020mes) in which the presence of doxycycline turns on ETV2 expression. After 1 week of doxycycline (1 μg ml^{-1}) induction of ETV2, cells were collected to determine the relative *ETV2* mRNA unit. Cells with a relative *ETV2* RNA unit within 60–100 were used for all experiments. An MOI of 50 was required for the inducible ETV2 lentiviral particles and rtTA lentiviral particles.

Lentivirus production

All lentiviral plasmids were prepared with a DNA Midiprep kit (Qiagen, 12145). Viruses were packaged in 293T cells by co-transduction with a second or third generation of packaging plasmids. Culture media were collected 48 h after transduction and virus particles were concentrated using a Lenti-X concentrator (Katara, 631232), resuspended in phosphate-buffered saline (PBS) without calcium or magnesium (Corning, 21040CV) and stored at −80 °C in small aliquots. Virus titres were determined with a Lenti-X p24 titre kit (Katara, 632200).

Tube-formation assays

Twenty-four-well plates were coated with 300 μl of Matrigel (Corning) for 30 min in a 37 °C incubator. Meanwhile, cells with or without ETV2 were accutased and counted. Cells were then resuspended in StemSpan (Stem Cell Technologies) supplemented with 10% knockout serum (Thermo Fisher Scientific, 10828028) and cytokines: 10 ng ml^{-1} FGF2 (bFGF) (Peprotech, 1000-18B), 10 ng ml^{-1} IGF1 (Peprotech, 100-11), 20 ng ml^{-1} EGF (Peprotech, AF-100-15), 20 ng ml^{-1} SCF (Peprotech, 300-07) and 10 ng ml^{-1} IL-6 (Peprotech, 200-06). One hundred thousand cells either with or without ETV2 were then dispersed in each well in 1 ml of medium. Cultures were placed in a 37 °C incubator with 5% oxygen for the remainder of the tube-formation experiments. The medium was changed every other day, by replacing 750 μl of medium with fresh medium. Care was taken to not disrupt the tubes during all medium changes. In several cases, a mixture of defined matrices comprising a mixture of laminin and entactin (Corning, 354259) and collagen IV (Corning, 354245) (LEC matrix) was used instead of Matrigel as indicated in the text. We combined these defined matrices at different ratios of laminin, entactin and collagen IV (LEC) components and ultimately found the most effective combination of these gel mixtures for tube-formation assays, which was: 200 μl of laminin and entactin (note that concentrations slightly vary for each lot; always diluted to 16.5 mg ml^{-1} in PBS first) and 100 μl of collagen IV (concentrations slightly vary for each lot; first diluted to 0.6 mg ml^{-1} in PBS), mixed together on the ice and stored at 4 °C overnight before use. The final format of the LEC matrix consisted of 11 mg ml^{-1} of the laminin and entactin mixture, and 0.2 mg ml^{-1} collagen IV. The volume of LEC was increased as needed, as long as the ratios and final concentrations were maintained. Vessel area was measured over the course of 24 h to 12 weeks for stage-2 (remodelling) and stage-3 (stabilization) phases. An EVOS inverted microscope with a 4× objective was used to capture images in their different (randomized) locations in each well for each condition and time point. All of the images were then analysed for the lumenized vessel area using ImageJ to trace the vessel area. The same procedure was used for cells transduced with ETS1 or myrAKT1[28], and for KRAS-transduced ECs[26].

Tube-formation assay in different medium formulations

ECs were accutased and plated on Matrigel at 100,000 cells per well in 24-well plates as described above. To assess the tube-formation assays of ETV2 ECs versus control ECs, we compared their capacity to form a tubular network in three different enriched pro-angiogenic medium formulations (Extended Data Fig. 1d). Medium formulation 1 (MF1) is a StemSpan tube-formation medium (Supplementary Data 2)—a serum-free medium containing StemSpan supplemented with knockout serum and cytokines (StemSpan (Stem Cell Technologies) supplemented with 10% knockout serum (Thermo Fisher Scientific, 10828028) and cytokines: 10 ng ml^{-1} FGF (Peprotech, 1000-18B), 10 ng ml^{-1} IGF1 (Peprotech, 100-11), 20 ng ml^{-1} EGF (Peprotech, AF-100-15), 20 ng ml^{-1} SCF (Peprotech, 300-07) and 10 ng ml^{-1} IL-6 (Peprotech, 200-06)). Medium formulation 2 (MF2, EGM-2) is an EC growth medium (Supplementary Data 2) (PromoCell, C22111). Medium formulation 3 (MF3) is the complete EC medium (Supplementary Data 2) with serum that was used to maintain and propagate ECs (400 ml

M199, 100 ml heat-inactivated FBS, 7.5 ml HEPES, 5 ml antibiotics (Thermo Fisher Scientific, 15070063), 5 ml glutamax (Thermo Fisher Scientific, 35050061), 5 ml lipid mixture (Thermo Fisher Scientific, 11905031) and 25 mg endothelial cell growth supplement (Alpha Aesar, J64516-MF). Media were changed every other day. Images were acquired at different time points. ImageJ was used to measure vessel area over time.

Video set-up for HUVECs cultured in 3D matrices in different medium formulations

GFP-labelled control HUVECs and R-VECs were embedded inside LEC matrix at 5 million cells per ml. Gels were polymerized on glass-bottomed culture dishes at 37 °C incubator for 15 min. Subsequently, either EGM-2 or StemSpan tube-formation medium (Supplementary Data 2) was added into the cell culture as described above. The medium was also supplemented with Trolox, a vitamin E analogue (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) (Sigma) at 100 μM to enable long-term imaging. The cultures were mounted in a temperature- and gas-controlled chamber for live-cell imaging. Time-lapse videos were acquired with a Zeiss Cell Observer confocal spinning disk microscope (Zeiss) equipped with a Photometrics Evolve 512 EMCCD camera at an interval of 40 min over 3 days. The medium was refreshed every two days.

Immunofluorescent staining of tubes in vitro

At 8 to 12 weeks all medium was removed from the wells. The tubes were washed once with PBS and fixed for 30 min in 4% paraformaldehyde (PFA) at room temperature. Then, the wells were rewashed with PBS and put in blocking buffer (containing 0.1% Triton-X) for 1 h at room temperature. For proliferation studies, a 16-h pulse of EdU (Click-iT EdU kit, Thermo Fisher Scientific, C10337) was used for all three stages of vessel formation.

Electron microscopy

Tissues were washed with serum-free medium or PBS then fixed with a modified Karmovsky's fix of 2.5% glutaraldehyde, 4% PFA and 0.02% picric acid in 0.1 M sodium cacodylate buffer at pH 7.2. After a secondary fixation in 1% osmium tetroxide and 1.5% potassium ferricyanide, samples were dehydrated through a graded ethanol series and embedded in an Epon analogue resin. Ultrathin sections were cut using a Diatome diamond knife (Diatome) on a Leica Ultracut S ultramicrotome (Leica). Sections were collected on copper grids, further contrasted with lead and viewed on a JEM 1400 electron microscope (JEOL) operated at 100 kV. Images were recorded with a Veleta 2k × 2k digital camera (Olympus SIS).

AFM measurements

AFM was used to examine the stiffness of HUVECs and adult human adipose ECs. Bright-field images of cells, for determination of the location of stiffness measurements, were acquired using an inverted microscope (Zeiss Axio Observer Z1) as the AFM base (20× 0.8 NA objective). An MFP-3D-BIO Atomic Force Microscope (Asylum Research) was used to collect force maps. A 5-μm borosilicate glass beaded probe (Novascan) with a nominal spring constant of 0.12 N m^{-1} was used for all measurements. Each force map sampled a 60 μm × 60-μm region, in a 20 × 20 grid of force curves (400 force curves total) under fluid conditions which covered an area of 360 μm^2. The trigger point was set to 2 nN with an approach velocity of 5 μm s^{-1}. The force-indentation curves were fit to the Hertz model for spherical tips using the Asylum Research software to determine Young's modulus, with an assumed Poisson's ratio value of 0.45 for the sample. Force maps of stiffness along with individual stiffness values for each measured point were then exported from the Asylum Research software for further analysis. A custom-made MATLAB (MathWorks) script was written to correctly analyse the data for the stiffness of the cells and filter measurements such that only data 1 μm from the glass bottom dish were analysed (to remove any substrate effect from the measurements).

RNA and protein collection from endothelial cell tubular capillaries

At indicated time points, capillaries of ECs from tube-formation assays were collected for RNA sequencing and western blotting. Before the cells were collected, the medium was completely removed from the well. Two millilitres of 2 mg ml^{-1} dispase (Roche 38621000) was added into each well to dissociate the EC tubes for 45 min at 37 °C with gentle shaking. Dissociated cells were pelleted, washed once in PBS and subsequently collected for either mRNA or protein isolation. On several occasions, dissociated ECs from tubes were pooled from multiple wells of the same EC line and experiment to allow sufficient isolation of mRNA and protein for downstream analysis.

Western immunoblot

Cells were lysed into 1× SDS loading buffer (50 mM Tris-HCl pH 6.8, 5% β-mercaptoethanol, 2% SDS, 0.01% bromophenol blue, 10% glycerol) followed by sonication (Bioruptor, 2 × 30 s at high setting). Proteins were solved on a 5–15% gradient Tris–glycine SDS–PAGE gel and semi-dry-transferred to nitrocellulose membranes. The following primary antibodies were used at the indicated dilutions: RAP1 (CST, 2399, 1:1,000); RASGRP3 (CST, 3334, 1:1,000), GAPDH (CST, 5174, 1:10,000); AKT (CST, 34685, 1:5,000); p-S473-AKT (CST, 4060, 1:2,000); ETS1 (CST, 14069, 1,000) and ETV2 (Abcam, ab181847, 1:1,000). All antibody information can be found in the Reporting Summary. Horseradish peroxidase (HRP)-conjugated secondary antibodies and the ECL prime western blotting system (GE Healthcare, RPN2232) were then used. Chemiluminescent signals were captured with a digital camera (Kindle Biosciences) and images of protein bands were taken for quantification using ImageJ.

In vivo experiments

All animal experiments were performed under the approval of the Weill Cornell Medicine Institutional Animal Care and Use Committee (IACUC). HUVECs transduced with an empty lentiviral vector or lentiviral vectors carrying ETV2 construct, and labelled with GFP or mCherry (2 million cells per plug), were injected subcutaneously into male or female 8–12-week-old SCID-beige mice (Taconic). The cells were first resuspended in PBS (50 μl) and then mixed with Matrigel (Corning, 356237) or LEC matrix as described above to a final volume of 350 μl. The gels were also supplied with FGF2 (10 ng ml^{-1}) (Peprotech, 1000-18B), VEGF-A (20 ng ml^{-1}) (Peprotech, 100-20) and heparin (100 μg ml^{-1}) (Sigma H3149-100KU). Each mouse received two plugs: one with control cells and the other with cells transduced with ETV2. Mice implanted with plugs were injected retro-orbitally with anti-human VEcad (clone BV9-Biolegend) conjugated to Alexa-647 (25 μg in 100 μl of PBS) or 70-kDa fluorescently labelled lysine-fixable dextran (Thermo Fisher Scientific) and euthanized 8 min after injection. Whole-mount images were taken directly on a Zeiss 710 confocal microscope using a well containing a coverslip bottom. The plugs were fixed in 4% PFA overnight and then dehydrated in ethanol or put in sucrose for further immunostaining. The dehydrated plugs were sent to Histoserv for further processing, sectioning and haematoxylin and eosin (H&E), picrosirius or Masson staining. The sections were processed for immunostaining as described below. GFP-labelled lentiviral KRAS-transduced cells were injected in mice as described above, but owing to a rapid increase in size, mice bearing plugs with KRAS-transduced cells were euthanized at 2 weeks.

Immunostaining of sections

Optimal cutting temperature compound (OCT) -frozen sections (20 μm), previously fixed in 4% PFA and treated in sucrose, were washed once with PBS. The slides were then incubated in blocking buffer (0.1% Triton-X, 5% normal donkey serum, 0.1% bovine serum albumin (BSA), for 30 min at room temperature and overnight in primary antibodies at the appropriate dilution (listed in the Reporting Summary) at 4 °C in blocking buffer. For thicker sections (50 µm) tissues were blocked overnight in blocking buffer at 4 °C (0.3% Triton-X, 5% normal donkey serum, 0.1% BSA) and then for two days in primary antibody in blocking buffer at 4 °C (0.3% Triton-X, 5% normal donkey serum, 0.1% BSA). The next day, the slides were washed 3 times for 10 min at room temperature and then incubated for three hours in fluorescently conjugated secondary antibodies (1:1,000). Finally, the slides were washed 3 times for 10 min and counterstained with DAPI. The sections were mounted on coverslips. A Zeiss 710 confocal or Zeiss Cell Observer confocal spinning disk microscope was used to acquire images. For stroma staining, a mouse anti-PDGFRβ antibody (1:500, Biolegend) or an anti-mouse SMA antibody (1:200, Abcam) was used. Mouse ECs were counterstained with mouse anti-endomucin antibody (1:100, Santa Cruz). Several images were taken from sections from different layers of each plug. At least 12 pictures (4 per mouse) from different slides were taken for each condition and time point. Images were processed using ImageJ and the percentage of vessel area within the area of each image field was quantified using the threshold feature in ImageJ.

RAP1 pull-down and western blots

A 10-cm plate of either HUVECs or ETV2-transduced HUVECs (flat 2D induction stage) was used for the active RAP1 assay (Cell Signaling, 8818S) according to the manufacturer's guidelines for the kit. In brief, the cells were washed once with PBS and then starved for three hours in M199 medium with 0.5% BSA. The cells were then scraped in the lysis buffer supplied with the kit and resuspended at around 1 mg ml^{-1}. A fraction was saved as input and the rest of the cells were used for RAP1–GTP pull-down. Positive and negative controls, as well as a beads-only control, were performed according to the manufacturer's guidelines. Proteins were solved on a 5–15% gradient Tris–glycine SDS–PAGE gel and semi-dry-transferred to nitrocellulose membranes. The membranes were then blocked in 5% milk in PBST and incubated in the provided RAP1 (1:1,000) antibody, GAPDH and/or ETV2 antibody for 48 h. After 48 h, the membranes were washed 3 times for 5 min and incubated in HRP-conjugated secondary antibody. Finally, after secondary washings, the membrane was blotted in ECL, chemiluminescent signals were captured with a digital camera (Kindle Biosciences) and images of protein bands were taken for densitometric quantification using ImageJ.

RAP1 inhibition experiment

Tube-formation assays for ECs with or without ETV2 were set up in 24 wells as described above. The next day, RAP1 inhibitor (GGTI-298, Tocris) resuspended in DMSO was added to the wells at a 1:1,000 dilution at the final concentration of 10 µM, and the same amount of DMSO was added to the control wells. The inhibitor and medium were changed every other day for 4 weeks. Images were obtained and the vessel area was calculated as described above at one-week and four-week time points.

RASGRP3 knockdown experiments

shERWOOD-UltramiR RASGRP3 shRNA lentiviral constructs (in pZIP-TRE3G) were purchased from TransOMIC Technologies. The clone number and targeted RASGRP3 sequences are as follows: ULTRA-3265848, AAGGGCAGAAGTCATCACAAA;ULTRA-3265850, CCTTGGAGTACACTTGAAAGA. The control shRNA (ULTRA-NT, ATGCTTTGCATACTTCTGCCT) targets a fly luciferase RNA sequence. Lentivirus was prepared as described above, using second-generation packaging plasmids. R-VECs (stage 1) were transduced with either shRNA virus or control shRNA virus (MOI = 3). Doxycycline was added at day 1 of the remodelling stage (stage 2) and the medium with doxycycline was replaced every other day for 4 weeks. Images were obtained and the vessel area was calculated as described above at two-week and four-week time points. To confirm *RASGRP3* knockdown, doxycycline was added to stage-1 R-VEC cells for one week and then the cells were collected for western blot analysis.

Proteasome inhibition experiment

R-VEC vessels were prepared on Matrigel as described above. At the stabilization stage (4 weeks), R-VEC tubes were treated with either 20 µM of MG132 (Selleck Chemicals) or DMSO for 6 h. The medium was removed and the wells were washed once with PBS. R-VEC tubes were then incubated in a solution of 2 mg ml^{-1} Dispase (Roche) for 45 min at 37 °C to dissociate the tubes. 20µM of MG132 (Selleck Chemicals) or DMSO was continuously provided during the dissociation period. Dissociated cells were collected and further processed for western blotting as described above.

Isolation of ECs from ETV2 reporter mice

ETV2-Venus reporter mice were a gift from V. Kouskoff[29]. In brief, embryos were isolated at E9.5 from pooled litters of ETV2-Venus reporter mice. For each independent biological replicate, five litters of mice at E9.5 were pooled together. All embryos were accutased for 20 min at 37 °C and then triturated several times with a pipette. The cells were post-stained for anti-mouse CD31 and anti-mouse CD45 antibodies, and then sorted as either ETV2$^{\text{Venus+}}$CD31^{+}CD45^{-} or CD31^{+}CD45^{-} (ARIAII, BD). Cells were sorted straight into Trizol-LS and the RNA was further purified using a Qiagen RNA-easy isolation kit.

Intestinal tissue collection and decellularization

Intestines were collected from Sprague Dawley rats ranging from 250–350 g in weight. In brief, under aseptic conditions a midline laparotomy was performed and the intestine exposed. A 5-cm-long intestinal segment was isolated, preserving the mesenteric artery and the mesenteric vein that perfuse the isolated segment. Both vessels were cannulated with a 26G cannula, and the intestinal lumen was cannulated using 1/4-inch barbed connectors. The isolated segments were decellularized, with perfusion through vasculature and lumen provided at 1 ml min^{-1} using a peristaltic pump (iPump). The decellularization process consisted of Milli-Q water for 24 h, sodium deoxycholate (Sigma) for 4h and DNase I (Sigma) for 3 h. Decellularized intestines were sterilized with gamma radiation before use.

Bioreactor culture

Decellularized intestines were seeded either with 5 million GFP^{+}ETV2^{+} human ECs or with 5 million GFP^{+} control ECs. Cells were seeded through the mesenteric artery and mesenteric vein. Seeded intestines were mounted inside a custom-made bioreactor under sterile conditions. After 24 h, perfusion was started through the mesenteric artery at 1 ml min^{-1} using a peristaltic pump (iPump). Cells were grown in complete EC medium (M199/EBSS (HyClone, SH302503.01) supplemented with 20% heat-inactivated FBS, 1% penicillin–streptomycin, 1.5% HEPES (Corning, 25-060-Cl), 1% glutamax (Gibco 35050-061), 1% lipid mixture (Gibco, 11905-031), 1% heparin (Sigma, H3149-100KU) and 15 µg ml^{-1} endothelial cell growth supplement (Merck, 324845)) for the first 5 days, and then cells were grown for 2 days in StemSpan (Stem Cell Technologies) supplemented with 10% knockout serum (Thermo Fisher Scientific, 10828028), 1% penicillin–streptomycin, 1% glutatamax, 10 ng ml^{-1} FGF2 (Peprotech100018B), 20 ng ml^{-1} EGF (InvitrogenPHG0311), 10ng ml^{-1} IGF2 (Peprotech 100-12), 20 ng ml^{-1} SCF (Peprotech 300-07) and 10 ng ml^{-1} IL-6 (Peprotech 200-06). After 7 days, re-endothelialized intestines were collected under sterile conditions and segments of 5 × 7 mm were excised for heterotopic implantation. The remaining intestinal tissue was then fixed in 4% PFA, mounted and prepared for imaging by fluorescent microscopy. To assess the patency of the vessels, some re-endothelialized intestines were perfused with fluorescently labelled LDL.

Heterotopic graft implantation

Mice used for these studies were maintained and experiments performed in accordance with the UK Animals (Scientific Procedures) Act 1986 and approved by the University College London Biological Services Ethical Review Process (PPL 70/7622). Animal husbandry at UCL Biological Services was in accordance with the UK Home Office certificate of designation. NOD-SCID-gamma (NSG) mice, aged between 8 and 12 weeks, were anaesthetized with a 2–5% isoflurane–oxygen gas mix for induction and maintenance. Buprenorphine (0.1 mg kg^{-1}) was administered at the induction of analgaesia. A midline laparotomy was performed under aseptic conditions. The stomach was externalized from the incision and the omentum stretched from the great curvature. A segment of the engineered intestine was then enveloped in the omentum, using 8/0 prolene sutures to secure the closure of the omental wrap. The stomach and the omentum were placed back in the abdomen and the laparotomy closed using 6/0 vicryl sutures. Mice were allowed to eat and drink normally immediately after surgery and no further medications were administered during the post-operative periods. After one week or four weeks, the mice were intravenously injected with fluorescently labelled anti-human VEcad (BV9 Biolegend) as described in 'In vivo experiments', and then euthanized. Grafts were retrieved together with the omental envelope and fixed in 4% PFA, mounted and prepared for imaging by fluorescent microscopy.

Analysis of vascular parameters for decellularized intestine experiments

Images for in vitro EC revascularization were processed using ImageJ by setting a threshold and quantifying the area covered by the CD31 signal with respect to the intestine area. In vivo quantification of cells positive for GFP and VEcad was performed on images acquired with a confocal microscope (Zeiss LSM710) and evaluation of vascular parameters was performed using Angiotool software (National Cancer Institute)[30].

Quantification of proliferating cells and apoptotic cells in decellularized scaffolds

Explanted intestinal grafts were fixed in 4% PFA, embedded in OCT and sectioned. Sections were stained for cleaved caspase 3 (Cell Signaling, 9661S) and for Ki67 (Abcam, AB15580). First, the sections were blocked for 1 h in PBS with 10% donkey serum. Then, primary antibodies were incubated overnight at 4 °C in blocking solution with the addition of 0.5% Triton-X. Primary antibodies were washed 3 times with PBS before the secondary antibody was added. Secondary antibody for donkey anti-mouse or rabbit (Alexa Fluor 547 or 647; Life Tech) was used at a dilution of 1:500 in blocking solution with 0.5% Triton X-100 and incubated at room temperature for 1 hour. Secondary antibody buffer was washed off with PBS 3 times and the slides mounted in a solution containing DAPI. Images were acquired with a confocal microscope (Zeiss LSM710). Three fields of view (425.10 μm × 425.10 μm in size) were evaluated per animal and the ratio between human VEcad (injected intra-vitally before euthanasia) and cleaved caspase 3- or Ki67-positive cells quantified.

Primary human pancreatic islets in static co-culture with ECs

Primary human islets were purchased from Prodo Laboratories. Twenty-five human islets were cultured alone, co-cultured with control ECs or co-cultured with R-VECs. Control ECs and R-VECs were used at 5 million cells per ml. The human islets with and without ECs were mixed in 40 μl Matrigel and plated into wells of a Nunc IVF 4-well dish (Thermo Fisher Scientific, 144444). Islets and ECs were co-cultured with serum-free islet medium (SFIM, Supplementary Data 2). The medium was composed of glucose-free RPMI 1640 supplemented with 0.1% human serum albumin, 10 μg ml^{-1} human transferrin, 50 μM ethanolamine, 50 μM phosphoethanolamine, 6.7 μg ml^{-1} sodium selenite, 10 ng ml^{-1} FGF2, 100 μg ml^{-1} heparin and 5.5 mM glucose. After two weeks of co-culture, samples were prepared for glucose-stimulated insulin secretion (GSIS). Samples were starved in Krebs-Ringer bicarbonate HEPES (KRBH) buffer containing 2 mM glucose for 2 h, followed by 45 min in 2 mM glucose as the basal insulin secretion and 45 min in 16.7 mM glucose as the stimulated insulin secretion. Insulin concentrations at the end of basal and stimulated phases were determined using the STELLUX Chemi Human Insulin ELISA (ALPCO). For each group, there were 11 replicates, with islets derived from 4 different donors. In other experiments, 200 human islets were cultured alone or mixed with 250,000 control ECs or 250,000 R-VECs in 50-μl Matrigel droplets. Human islet explants in co-culture were stained for EpCAM and VEcad and imaged at one and two weeks. In brief, the growth medium was removed and the cells were fixed in 4% PFA for 20 min. They were then permeabilized in 0.5% Triton-X for 20 min and blocked in IF Buffer (PBS, 0.2% Triton-X, 0.05% Tween, 1% BSA) for 1 h. Then, the cells were incubated in primary antibodies overnight in IF buffer: anti-EpCAM (1:100, Biolegend), VEcad (1:100, R&D). They were then washed 3 times with PBS 0.1% Tween. The wells were then incubated with secondary antibodies (1:1,000) in the IF buffer for 3 h. The solution was removed, DAPI in PBS was added for 5 min and cells were washed twice with PBS 0.1% Tween.

To quantify the interacting vessels with human pancreatic islets, co-cultures were imaged using a 10× objective to capture both GFP-labelled vessels and human pancreatic islets in the bright field. Using the custom MATLAB code, we traced the area of GFP-labelled vessels that surrounded and wrapped the human pancreatic islets for co-cultures with control ECs and with R-VECs.

Vascular network formation in microfluidic devices

We produced a more substantial scale device using photo-lithography as previously described[17]. The distance between the two fluidic channels or the width of the device is 3 mm (increased from 1 mm). The length of the device or the length of the fluidic channels is 5 mm. The height of the device is 1 mm. The total volume of the device is 15 μl. In brief, each device comprises two layers of poly(dimethylsiloxane) (PDMS; Sylgard 184; Dow-Corning), which are cast from silicon wafer masters. The devices are plasma-treated with plasma etcher (Plasma Etch) and subsequently treated with (3-glycidyloxypropyl)trimethoxysilane (Sigma, 440167) overnight. The next day, they are submerged in water to wash overnight before use. All devices are kept in a 37 °C incubator with 20% oxygen.

A mixture of 3 million ml^{-1} ETV2 HUVECs or control HUVECs in 5 mg ml^{-1} bovine fibrinogen (Sigma) and 3 U ml^{-1} bovine thrombin (Sigma) was injected into the devices with two 400-μm acupuncture needles (Hwato). After the cell and gel mixture polymerized, the acupuncture needles were pulled out leaving two hollow channels. HUVECs were seeded into the hollow channels to form two parent vessels on the next day. The devices were placed on a platform rocker for the entire experiment (Benchmark). Cells were cultured in the medium for vessel formation in microfluidic devices (Supplementary Data 2) and refreshed daily until day 7, when the devices were fixed and imaged.

For human pancreatic islet culture experiments, devices were set up similarly to experiments with ECs alone. Approximately 75 human pancreatic islets were mixed either alone or with control ECs or R-VECs (4 million cells per ml) cells in 5 mg ml^{-1} bovine fibrinogen and 3 U ml^{-1} bovine thrombin to a total volume of 30 μl and injected into the devices. The needles were removed after fibrin gel polymerization, and 200 μl of human pancreatic islet co-culture medium (Supplementary Data 2) was added into each of the fluidic channels. The devices were placed on a platform rocker (Benchmark 2000) during the entire experiment.

GSIS assay for human pancreatic islets in the devices

Human pancreatic islets were placed in the devices as described above either alone or in co-culture with control ECs or R-VECs. Human cadaveric islets (Prodo Labs) were procured from three healthy separate donors, with a total of $n = 4$ devices for no ECs, $n = 4$ devices for control

ECs and $n = 8$ devices for R-VECs. After 4 days, the medium was removed in all the devices. The devices were then starved with 2 mM glucose for 2 h in the incubator. At the end of starvation, 300 µl of 2 mM glucose KRBH buffer was added at the inlet of the device, and devices were incubated at 37 °C for 3 min. Driven by gravity, KRBH buffer perfused through to the other side (outlet) of the device during the incubation. After the 3-min incubation, fluid from the outlets was collected for insulin measurement through ELISA. The inlets were also emptied of any remaining fluid. Then, another 300 µl KRBH buffer was added to inlets, leaving the outlets empty. In R-VEC co-culture devices, 30–150 µl fluid collected in the outlets, owing to high perfusion rates. In islets-alone and control-EC co-culture devices, only a small amount of fluid (less than 10 µl) was found in the outlets. To enable sample collection, we rinsed the outlets of islets alone and control-EC co-culture devices with 150 µl KRBH buffer and collected all outlet liquid for insulin measurement using ELISA. Sample collection was repeated for a total of 8 times using 2 mM glucose KRBH buffer, and another 8 times using 16.7 mM glucose KRBH buffer. In the end, we acquired a series of semi-dynamic GSIS samples. We examined the insulin concentration at the outlet of the device at the third (at $t = 9$ min) and eighth ($t = 24$ min) collections at both the 2 mM and the 16.7 mM glucose phases. The insulin level per device was calculated as: insulin per device = insulin concentration × collected volume. Basal insulin levels were determined as the average of the third and eighth collections at 2 mM glucose. Insulin concentration was determined using the STELLUX Chemi Human Insulin ELISA (ALPCO).

Staining protocol for experiments in devices

To stain for ECs in the devices, immediately before the experiment was terminated, all medium was aspirated in both fluidic channels in the devices. VEcad antibody (200 µl) conjugated with Alexa 647 at 10 µg ml^{-1} (Biolegend) was placed in one of the fluidic channels and allowed to slowly perfuse through the lumenized R-VEC vessels for 15–20 min in the incubator from one fluidic channel to the other fluidic channel. The device was then washed 3 times with basal medium and fixed with PFA for 45 min.

When co-culture experiments were set up with human pancreatic islets, the same protocol was used to stain for R-VEC lumenized vessels with VEcad-conjugated antibody. Post-fixation, the device was permeabilized with 0.1% Triton-X for 45 min and further stained with either EpCAM for human COs or EpCAM and insulin for human pancreatic islets. To stain for EpCAM (Biolegend) the conjugated antibodies were added to both fluidic channels at 10 µg ml^{-1} for 48 h on a rocker at 4 °C. The devices were washed 3 times with 1× PBS and subsequently washed and submerged into 1× PBS for 24 h on a rocker at 4 °C. A similar staining procedure was used for insulin and post-VEcad staining, except that permeabilization was carried out overnight, followed by primary antibody staining as described above and secondary staining for 24 h on a rocker at 4 °C. The devices went through washing for another 24 h with 1× PBS on a rocker at 4 °C and were then imaged using a Zeiss 710 confocal microscope.

Whole-blood perfusion in vascularized microfluidic devices

For blood perfusion videos, vessels were prepared with 3 million R-VEC cells per ml, as described above. Medium (400 µl) (Promocell) was refreshed. On day 7, blood was collected from a donor following IRB protocol in a heparinized tube. We sealed one end of both of the fluidic channels leaving two reservoirs diagonal to one another open for the perfusion experiment. Human heparinized whole peripheral blood (BD Vacutainer) was obtained from consented healthy individuals by phlebotomy. Then 200 microlitres of whole blood were immediately pipetted into one of the fluidic channels at the open reservoir. The blood cells along with intact plasma entered the fluidic channel, traversed through the lumenized R-VEC vessels and exited to the reservoir diagonal to the reservoir in which blood entered. In experiments to perfuse blood in devices with R-VECs in co-culture with human pancreatic islets, we stained blood cells with PKH26 red fluorescent dye (Sigma, MMIDI26-1KT) according to the manufacturer's protocol for 5 min on ice. Fluorescently labelled blood cells were pipetted into the reservoir, traversed through the lumenized R-VEC vessels and exited to the diagonal reservoir. In other devices (control ECs + human pancreatic islets, and human pancreatic islets alone), fluorescently labelled blood cells were not able to traverse from one fluidic channel to the other fluidic channel. Images were taken with an Axio Observer Z1 equipped with Hamamatsu Flash 4.0 v2, sCMOS camera and 10×/0.45 objective.

Isolation and culture of mouse small intestine organoids

Mouse small intestine organoids were isolated as previously described[31]. Fifteen centimetres of the proximal small intestine was removed and flushed with cold PBS. After opening longitudinally, it was washed in cold PBS until the supernatant was clear. The intestine was then cut into 5-mm pieces and placed into 10 ml cold 5 mM EDTA-PBS and vigorously resuspended using a 10-ml pipette. The supernatant was aspirated and replaced with 10 ml EDTA and placed at 4 °C on a benchtop roller for 10 min. This was then repeated for a second time for 30 min. The supernatant was aspirated and then 10 ml of cold PBS was added to the intestine and resuspended with a 10 ml pipette. After collecting this 10 ml fraction of PBS containing crypts, this was repeated and each successive fraction was collected and examined underneath the microscope for the presence of intact intestinal crypts and lack of villi. The 10-ml fraction was then mixed with 10 ml DMEM basal medium (Advanced DMEM F12 containing penicillin–streptomycin, glutamine, HEPES (10 mM) and 1mM *N*-acetylcysteine (Sigma Aldrich A9165-SG) containing 10 U ml^{-1} DNase I (Roche, 04716728001), and filtered through a 100-µm filter into a BSA (1%)-coated tube. It was then filtered through a 70-µm filter into a BSA (1%)-coated tube and spun at 1,200 rpm for 3 min. The supernatant was aspirated and the cell pellet mixed with 5 ml basal medium containing 5% FBS and centrifuged at 200*g* for 5 min. The purified crypts were then resuspended in basal medium and mixed 1:10 with Growth Factor Reduced (GFR) Matrigel (Corning, 354230). A 40-µl sample of the resuspension fluid was plated in a 48 well plate and allowed to polymerize. Mouse small intestine organoid growth medium composed of basal medium containing 40 ng ml^{-1} EGF (Invitrogen PMG8043), 100 ng ml^{-1} Noggin (Peprotech 250-38) and 500 ng ml^{-1} R-spondin1 (R&D Systems, 3474-RS-050), were then laid on top of the Matrigel. In some experiments, small intestinal organoid growth medium was made with R-spondin1 from conditioned medium, collected from HEK293 cell lines expressing recombinant R-spondin1 (provided by C. Kuo).

Maintenance of mouse small intestine organoids

The medium of the organoids was changed every two days and they were passaged 1:4 every 5–7 days. To passage, the growth medium was removed and the Matrigel was resuspended in cold PBS and transferred to a 15-ml falcon tube. The organoids were mechanically disassociated using a p1000 or a p200 pipette and pipetting 50–100 times. Seven ml of cold PBS was added to the tube and pipetted 20 times to fully wash the cells. The cells were then centrifuged at 1,000 rpm for 5 min and the supernatant was aspirated. They were then resuspended in GFR Matrigel and replated as above. For freezing, after spinning the cells were resuspended in basal medium containing 10% FBS and 10% DMSO and stored in liquid nitrogen indefinitely.

Mouse small intestine organoid co-culture and staining

Mouse small intestine organoids were co-cultured for 4–7 days either alone or with control ECs or R-VECs, at a final concentration of 5 million cell per ml of Matrigel. Organoids were mechanically dissociated as described above and mixed with the ECs, spun down and resuspended in GFR Matrigel. The mixture was then dispersed in 30-µl droplets in 8-well chamber slides (Lab-Tek II, 154534) or 50-µl droplets in a Nunc

IVF 4-well dish (Thermo Fisher Scientific, 144444). Cells were cultured in mouse small intestine organoid medium (Supplementary Data 2) as described above (EGF 40 ng ml^{-1}, Noggin 50 ng ml^{-1}, R-spondin1 conditioned medium (10%) + FGF-2 (10 ng ml^{-1}) (Peprotech, 1000-18B) and heparin (100 µg ml^{-1}) (Sigma, H3149-100KU). Vessel area was quantified by the threshold function in ImageJ and individual sprouts in contact with the mouse small intestine organoids were counted and reported as vessel sprouts per organoid. Where indicated, 10 µM EdU was added to the growth medium for 6 h before fixing. The growth medium was removed and the cells were fixed in 4% PFA for 20 min. They were then permeabilized in 0.5% Triton-X for 20 min and blocked in IF buffer (PBS, 0.2% Triton-X, 0.05% Tween, 1% BSA) for 1 h or immediately processed for EdU staining according to directions provided with the Click-iT Edu Imaging Kit (Invitrogen C10340). For immunofluorescent staining, cells were incubated in primary antibodies overnight in IF buffer: anti-KRT20 (1:200, Cell Signaling Technologies, 13063). They were then washed 3 times with PBS 0.1% Tween. The wells were then incubated with secondary antibodies (1:1,000) in the IF buffer for 3 h. The solution was removed, DAPI in PBS was added for 5 min and cells were washed twice with PBS 0.1% Tween. The chambers were then removed and cover slips were mounted using Prolong Gold antifade medium (Invitrogen P36930).

Isolation and culture of human normal and tumour colon organoids

Isolation of human colonic crypts and adenoma, and culture and maintenance of organoid cultures, were performed as previously described[32]. Normal and adenoma tissues were collected from colonic resections according to protocols approved by the Weill Cornell Medicine IRB. In brief, human colonic mucosa samples were obtained by trimming surgically resected specimens. The underlying muscle layer was removed using fine scissors under a stereomicroscope, leaving the mucosa, which was cut into 5-mm pieces on a Petri dish, placed into a 15-ml centrifuge tube containing 10 ml of cold DPBS and washed 3 times. Ten millilitres of cold DPBS supplemented with 2.5 mM EDTA was added to the tube and the tube was incubated for 1 h at room temperature with gentle shaking. Isolated crypts were mixed with Matrigel (Corning, 354230), dispensed in the centre of each well of a 6-well plate using a 200-µl pipette and placed at 37 °C for 10 min to solidify the Matrigel.

Normal COs were procured from Jason Spence's (J.S.) laboratory at the University of Michigan as previously described[33,34] (specifically, human CO lines 87 and 89). Healthy human COs were passaged 1:3 every 7 days by mechanical dissociation (pipetting) and grown in 12-well low-attachment plates in 30-µl Matrigel droplets. Normal COs were cultured in human CO medium (Supplementary Data 2) comprising Advanced DMEM F12, penicillin–streptomycin, 4 mM glutamax, 1% HEPES, primocin (100 µg ml^{-1}), 50% L-WRN (WNT3a, R-spondin1, Noggin)-conditioned medium, N2, B27 without vitamin A, *N*-acetylcysteine (1 mM), human recombinant EGF (50 ng ml^{-1}), Y-27632 (10 µM), A-83-01 (500 nM) and SB202190 (10 µM). The L-WRN-conditioned medium was generated using L-WRN cells. Conditioned medium was collected for 4 days, pooled, sterile-filtered and frozen into aliquots until use.

Human CRCOs were procured through the Institute for Precision Medicine at Weill Cornell Medicine[35]. The CRCOs were split 1:3 every 7 days by digesting in TrypLE Select (Thermo Fisher Scientific) supplemented with 10 µM Y27632 (Tocris Bioscience), and were maintained in human CRCO medium and propagated in GFR Matrigel. Human CRCO medium (Supplementary Data 2) comprises Advanced DMEM F12, 1% penicillin–streptomycin, 1% glutamax, 1% HEPES, R-spondin1-conditioned medium (5%), *N*-acetylcysteine (1.25 mM), human recombinant EGF (50 ng ml^{-1}), human recombinant FGF-10 (20 ng ml^{-1}), FGF-2 (1 ng ml^{-1}), Y-27632 (10 µM), A-83-01 (500 nM), SB202190 (10 µM), nicotinamide (10 mM), PGE2 (1 µM), NRG (10 ng ml^{-1}) and human gastrin 1 (10 nM).

Co-cultures of normal and tumour organoids with ECs

R-VECs or control ECs (at a final concentration of 5 million cells per ml) were mixed with healthy human COs or patient-derived CRCOs, spun down and resuspended in Matrigel (Corning, 354230) or LEC matrix as described above. The cells were then dispersed in 30–70-µl Matrigel or LEC droplets in 8-well chamber slides (Lab-Tek II, 154534) or a Nunc IVF 4-well dish (Thermo Fisher Scientific, 144444) and cultured in the respective organoid medium with the addition of FGF-2 (10 ng ml^{-1}) (Peprotech, 1000-18B) and heparin (100 µg ml^{-1}) (Sigma H3149-100KU). The medium was changed every other day. A 4.5-h pulse of EdU was used for all tumour organoid co-culture experiments (Click-iT EdU kit, Invitrogen C10340). The co-cultures were maintained in a 37 °C incubator with 20% oxygen. Human COs and CRCOs were stained similarly to mouse small intestinal organoid co-cultures. Antibodies against human EpCAM (Biolegend) and VEcad (R&D) were added and co-cultures were incubated overnight, followed by secondary antibody staining.

Preparation of normal and tumour organoids cultured with ECs for molecular profiling and single-cell sequencing

For single-cell sequencing, co-cultures were maintained for seven days. To collect cells in co-culture for single-cell sequencing, the medium was removed from the culture and the organoid–endothelial cell droplets were incubated in 2 mg ml^{-1} dispase (Roche) for 20 min at 37 °C with shaking. The cells were then spun down and incubated for an additional 15 min at 37 °C in accutase. At this point, the endothelial cells were mostly released from the co-cultures and collected by filtering through a 40-µm mesh. The rest of the undigested cells (mainly organoid clusters) were further dissociated into single cells by incubating with TrypIE for an additional 45 min at 37 °C until the cells were completely separated as single cells. This two-step digestion allowed for increased viability and efficient dissociation of both endothelial cells and organoids. Both the first and the second fraction were further processed for single-cell analysis. Single cells were collected and filtered through a 35-µm nylon mesh and processed for single-cell sequencing.

For qRT–PCR experiments, co-cultures were maintained for seven days in Matrigel. To collect cells and dissociate organoids in co-cultures, we incubated the Matrigel droplets with TrypLE-Express enzyme (Thermo Fisher Scientific, 3 ml per 50-µl Matrigel droplet) for 45 min at 37 °C with vigorous shaking. The dissociated cells were then washed twice, once with organoid culture medium and once with MACs buffer. Dissociated cells were resuspended in 100 µl MACS buffer and anti-human CD31 (Biolegend, 10 µg ml^{-1}) was used to stain for endothelial cells for 30 min on ice. The cell suspension was washed with MACS buffer and resuspended in MACS buffer with DAPI (1µg ml^{-1}). Subsequently, cells were sorted to purify the DAPI^{-}CD31^{-} population. An Accurus PicoPure RNA isolation kit (Thermo Fisher Scientific) was used to isolate RNA from the collected cells.

Quantification of vessels that interact with human COs and CRCOs in serial confocal videos

Human CRCOs and COs were stained with CellTracker (Invitrogen, C34565) as per the instruction manual of the manufacturer. CRCOs and COs were embedded inside Matrigel or LEC matrix with either control ECs or R-VECs at 5 million cells per ml. A mixture of gel and cells was pipetted onto a glass-bottomed dish and polymerized inside a 37 °C incubator for 15 min. The culture was then fed with organoid medium supplemented with 10 ng ml^{-1} bFGF (Peprotech) and 100 µg ml^{-1} heparin (Sigma H3149-100KU). To enable long-term imaging, 6-hydroxy-2,5,7,8 -tetramethylchroman-2-carboxylic acid (Sigma), as an antioxidant, was also added into the medium at 100 µM. The culture was immediately mounted onto a temperature- and gas-controlled chamber. Time-lapse videos were acquired with a Zeiss Cell Observer confocal spinning disk microscope (Zeiss) equipped with a Photometrics Evolve 512 EMCCD camera at an interval of 40 min over 3–4 days. Medium was refreshed every two days.

To quantify the vessels interacting with normal and tumour colon organoids, *Z*-projection images of time-lapse videos from several time points were obtained using ImageJ. Custom MATLAB codes were written to quantify the interacting vessel areas with all individual organoids. The custom MATLAB codes are provided in Supplementary Data 3. In brief, the code was used to manually trace the perimeter of all vessels around which ECs were wrapping and tapping the organoids. The area of the manually traced interacting vessels was quantified and reported.

RNA library preparation and analysis of sequencing data

RNA was isolated and purified using the Rneasy Mini Kit (Qiagen) or Accurus PicoPure RNA isolation kit (Thermo Fisher Scientific). RNA quality was verified using an Agilent Technologies 2100 Bioanalyzer. RNA libraries were prepared and multiplexed using the Illumina TruSeq RNA Library Preparation Kit v.2 (non-stranded and poly-A selection) and 10 nM of cDNA was used as input for high-throughput sequencing with Illumina's HiSeq 2500 or HiSeq 4000, producing 51-bp paired-end reads. Sequencing reads were de-multiplexed (bcl2fastq) and mapped with STAR v.2.6.0c (ref. [36]) with default parameters to the appropriate NCBI reference genome (GRCh38.p12 for human samples and GRCm38.p6 for mouse samples). Fragments per gene were counted with featureCounts v.1.6.2 (ref. [37]) with respect to Gencode comprehensive gene annotations (release 28 for human samples and M17 for mouse samples).

Transcriptome data analysis

Differential gene expression analysis was performed using DESeq2 v.1.18.1 (ref. [38]), and only false discovery rate (FDR)-adjusted *P* values of less than 0.05 were considered statistically significant. Before differential gene expression analysis, genes expressed at low levels were filtered out by only retaining genes that have more than 1CPM in the condition with the least number of replicates. Base-2 log-transformed CPM values were used for heat map plots, which were centred and scaled by row. Before visualization, tissue-specific effects were removed using the removeBatchEffect function from limma v.3.34.9 (ref. [39]). GO analysis was performed using DAVID Bioinformatics Resource Tools v.6.8 (ref. [40]).

ChIP and antibodies

To identify the genome-wide localization of ETV2, K4me3, K27me3 and K27ac modifications in R-VECs or control ECs, ChIP assays were performed with approximately 1×10^7 cells per experiment, as previously described[41]. Cells introduced with triple Flag-tagged ETV2 lentivirus (as described above) were used for the ETV2 ChIP. In brief, cells were cross-linked in 1%PFA for 10 min at 37 °C, then quenched by 0.125 M glycine. Chromatin was sheared using a Bioruptor (Diagenode) to create fragments of 200–400 bp, immunoprecipitated by 2–5 μg of antibody or mouse IgG bound to 75 μl Dynabeads M-280 (Invitrogen) and incubated overnight at 4 °C. Magnetic beads were washed and chromatin was eluted. The ChIP DNA was reverse-cross-linked and column-purified. All ChIP antibodies are identified in the Reporting Summary.

ChIP–seq library construction and sequencing

ChIP–seq libraries were prepared with the Illumina TruSeq ChIP Library Preparation Kit for DNA from ETV2 ChIP, and K4me3, K27me3 and K27ac modification ChIP. ChIP–seq libraries were sequenced with the Illumina HiSeq 4000 system.

ChIP–seq data processing and analysis

ChIP–seq reads were aligned to the reference human genome (hg19, GRCh37) using the BWA alignment software (v.0.5.9)[42]. Unique reads mapped to a single best-matching location with no more than 4% of the read length of mismatches were kept for peak identification and profile generation. Sequence data were visualized with IGV by normalizing to 1 million reads[43]. The software MACS2 (ref. [44]) was applied to the ChIP–seq data with sequencing data from input DNA as control to identify genomic enrichment (peak) of ETV2. The SICER (v.1.1) (ref. [45]) algorithm was applied to the ChIP–seq data with sequencing data from input DNA as a control to identify genomic regions with significant enrichment differences in different cell types. The resulting peaks were filtered by $P < 0.05$ for ETV2 and FDR < 0.01 for histone modifications. We computed the read counts in individual promoters using HOMER[46]. Each identified peak was annotated to promoters (±2 kb from transcription start site), gene body or intergenic region by HOMER. Summary and peak call information for all ChIP–seq data processing and analysis is provided in Supplementary Table 1.

10X Chromium single-cell transcriptomics and analysis

The following two experiments were performed for single-cell library preparation to establish the adaptation of R-VECs when co-cultured with normal or malignant organoids.

Experiment 1: R-VECs were co-cultured alone or together with human COs for seven days in human CO medium supplemented with 10 ng ml^{-1} FGF2 (Promocell) and 100 μg ml^{-1} heparin. COs were also cultured alone in CO medium supplemented with 10 ng ml^{-1} FGF and 100 ng ml^{-1} heparin for seven days. After seven days, all three conditions (R-VECs alone, R-VECs + COs, or COs alone) were dissociated with dispase and TryplE (Thermo Fisher Scientific) as described above, and submitted for 10X Chromium single-cell analysis. All three samples were processed and run at the same time.

Experiment 2: R-VECs were co-cultured alone or together with human CRCOs for seven days in CRCO medium supplemented with 10 ng ml^{-1} FGF2 and 100 μg ml^{-1} heparin. The CRCOs were also cultured alone in CRCO medium with 10 ng ml^{-1} FGF and 100 μg ml^{-1} heparin for 7 days. After 7 days, all three conditions (R-VECs alone, R-VECs + CRCOs, or CRCOs alone) were dissociated with collagenase, dispase and TryplE as described above, and submitted for 10 Chromium single-cell analysis. All three samples were processed and run at the same time.

The single-cell suspension was loaded onto a well on a 10X Chromium Single Cell instrument (10X Genomics). Barcoding and cDNA synthesis were performed according to the manufacturer's instructions. In brief, the 10X GemCode Technology partitions thousands of cells into nanolitre-scale gel bead-in-emulsions (GEMs), in which all the cDNA generated from an individual cell share a common 10X barcode. To identify the PCR duplicates, a unique molecular identifier (UMI) was also added. The GEMs were incubated with enzymes to produce full length cDNA, which was then amplified by PCR to generate enough quantity for library construction. Qualitative analysis was performed using the Agilent Bioanalyzer High Sensitivity assay. The cDNA libraries were constructed using the 10X Chromium single-cell 3′ Library Kit according to the manufacturer's original protocol. In brief, after the cDNA amplification, enzymatic fragmentation and size selection were performed using SPRI select reagent (Beckman Coulter, B23317) to optimize the cDNA size. P5, P7, a sample index and read 2 (R2) primer sequence were added by end repair, A-tailing, adaptor ligation and sample-index PCR. The final single-cell 3′ library contains standard Illumina paired-end constructs (P5 and P7), Read 1 (R1) primer sequence, 16-bp 10X barcode, 10-bp randomer, 98-bp cDNA fragments, R2 primer sequence and 8-bp sample index. For quality control after library construction, 1 μl of the sample was diluted 1:10 and run on the Agilent Bioanalyzer High Sensitivity chip for qualitative analysis. For quantification, the Illumina Library Quantification Kit (KAPA Biosystems, KK4824) was used.

Libraries were sequenced on an Illumina NextSeq500 with a 150-cycle kit using the following read length: 26-bp Read 1 for cell barcode and UMI, 8-bp I7 index for sample index and 132-bp Read 2 for transcript. Cell Ranger 2.2.0 (http://10xgenomics.com) was used to process Chromium single-cell 3′ RNA-seq output. First, ‘cellranger mkfastq’ demultiplexed the sequencing samples based on the 8-bp sample index read to generate fastq files for the Read 1 and Read 2, followed by extraction of 16-bp cell barcode and 10-bp UMI. Second, ‘cellranger count’ aligned the Read 2 to the human reference genome (GRCh38) using STAR[36]. Then aligned reads were used to generate the data matrix only when they have valid barcodes and UMI and map to exons (Ensembl GRCh38) without PCR duplicates. Valid cell barcodes were defined on the basis of UMI distribution.

All single-cell analyses were performed using the Seurat package in R (v.2.3.4) (ref. [47]). Once the gene–cell data matrix was generated, poor-quality cells were excluded, including cells with more than 6,000 uniquely expressed genes (as they are potentially cell doublets). Only genes expressed in three or more cells in a sample were used for further analysis. Cells were also discarded if their mitochondrial gene percentages were over 10% or if they expressed fewer than 600 unique genes, resulting in 20,778 genes across 24,478 cells, with the median UMI count for each cell across the entire dataset being 7,845 and the median number of unique genes per cell being 2,397. Further information on each sample that passed the quality filters is available in Supplementary Table 2. Following best practices in the package suggestions, UMI counts were log-normalized and after the most highly variable genes were selected the data matrices were scaled using a linear model, with variation arising from UMI counts and mitochondrial gene expression mitigated for. Principal component analysis was subsequently performed on this matrix and after reviewing principal component heat maps and jackstraw plots, UMAP visualization was performed on the top 29 components and the clustering resolution was set at 1.0 for visualizations. Differential gene expression for gene-marker discovery across the clusters was performed using the Wilcoxon rank-sum test in the Seurat package.

Epithelial cells were identified by the epithelial cell markers *EPCAM*, *CDH1* and *KRT19* and ECs were identified by the EC markers *VEcad*, *CD31* and *VEGFR2*. Epithelial cells were filtered out from the next analysis to identify heterogeneity amongst the EC populations of the co-cultured normal and tumour cell populations. The epithelial cell fraction was also analysed on its own in the tumour and co-cultured samples. In both of these analyses best practices were again followed for cluster discovery using the top 20 components and cluster resolution 0.6 in the matched tumour and normal sample sets and differential gene expression for gene-marker discovery across the clusters were performed using the Wilcoxon rank-sum test in the Seurat package.

Statistical analysis and data reporting

Data were assessed and analysed using appropriate statistical methods. The normality of data was assessed using the Kolmogorov–Smirnov test. Sample sizes and statistics for each experiment are provided in Supplementary Data 1. GraphPad Prism v.7 was used for all statistical analysis, unless otherwise indicated. No statistical methods were used to determine sample size. Unless otherwise stated, the experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Source ChIP–seq data are provided in Supplementary Table 1 and source scRNA-seq data are provided in Supplementary Table 2. The RNA-seq data can be viewed at the Gene Expression Omnibus (GEO) under accession number GSE131039. The ChIP–seq data and scRNA-seq data can be viewed at the GEO under accession numbers GSE147746 and GSE148996, respectively. Source data are provided with this paper.

Code availability

All of the code used in this paper is available from the authors on request.

25. Baudin, B., Bruneel, A., Bosselut, N. & Vaubourdolle, M. A protocol for isolation and culture of human umbilical vein endothelial cells. *Nat. Protocols* **2**, 481–485 (2007).
26. Seandel, M. et al. Generation of a functional and durable vascular niche by the adenoviral *E4ORF1* gene. *Proc. Natl Acad. Sci. USA* **105**, 19288–19293 (2008).
27. Ginsberg, M., Schachterle, W., Shido, K. & Rafii, S. Direct conversion of human amniotic cells into endothelial cells without transitioning through a pluripotent state. *Nat. Protocols* **10**, 1975–1985 (2015).
28. Schachterle, W. et al. Sox17 drives functional engraftment of endothelium converted from non-vascular cells. *Nat. Commun.* **8**, 13963 (2017).
29. Wareing, S., Eliades, A., Lacaud, G. & Kouskoff, V. ETV2 expression marks blood and endothelium precursors, including hemogenic endothelium, at the onset of blood development. *Dev. Dyn.* **241**, 1454–1464 (2012).
30. Zudaire, E., Gambardella, L., Kurcz, C. & Vermeren, S. A computational tool for quantitative analysis of vascular networks. *PLoS One* **6**, e27385 (2011).
31. Sato, T. et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
32. Sugimoto, S. & Sato, T. Establishment of 3D intestinal organoid cultures from intestinal stem cells. *Methods Mol. Biol.* **1612**, 97–105 (2017).
33. Dame, M. K. et al. Identification, isolation and characterization of human LGR5-positive colon adenoma cells. *Development* **145**, dev153049 (2018).
34. Tsai, Y. H. et al. A method for cryogenic preservation of human biopsy specimens and subsequent organoid culture. *Cell. Mol. Gastroenterol. Hepatol.* **6**, 218–222.e7 (2018).
35. Puca, L. et al. Patient derived organoids to model rare prostate cancer phenotypes. *Nat. Commun.* **9**, 2404 (2018).
36. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
37. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
38. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
39. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
40. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4**, 44–57 (2009).
41. Liu, Y. et al. Epigenetic profiles signify cell fate plasticity in unipotent spermatogonial stem and progenitor cells. *Nat. Commun.* **7**, 11275 (2016).
42. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
43. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
44. Zhang, Y. et al. Model-based analysis of ChIP–seq (MACS). *Genome Biol.* **9**, R137 (2008).
45. Zang, C. et al. A clustering approach for identification of enriched domains from histone modification ChIP–seq data. *Bioinformatics* **25**, 1952–1958 (2009).
46. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
47. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

Acknowledgements S.R. is supported by the Ansary Stem Cell Institute; grants from the National Institutes of Health (NIH) (R35 HL150809, R01s DK095039, HL119872, HL128158, HL115128, HL139056, RC2 DK114777 and U01AI138329); the Empire State Stem Cell Board; grants from New York State Stem Cell Science (NYSTEM) (C026878, C028117, C029156, C030160); the Daedalus Fund for Innovation from Weill Cornell Medicine; and the Starr Foundation stem cell core project; and Qatar National Priorities Research Program (NPRP 8-1898-3-392); and the Tri-Institutional Stem Cell Initiatives (TRI-SCI 2013-032, 2014-023, 2016-013 and 2019-029). R.E.S. is supported by NIH grants R01CA234614, 2R01AI107301, R01DK121072,1R21AI117213, RO3DK117252, R01AA027327 and K08DK101754 and Department of Defense CA170574. R.E.S. is a Irma Hirschl Trust Research Award Scholars. R.S. is supported by NYSTEM contract C32596GG; P.D.C. by NIHR Research (NIHR-RP-2014-04-046); P.D.C., A.F.P., A.M.T. and F.S. by the OAK Foundation (W1095/OCAY-14-191), H2020 grant INTENS 668294 and the NIHR Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust; A.M.T. by the BBSRC ICASE studentship 167881; Y. Liu and J.M.G.-S. by a New York Stem Cell Foundation Druckenmiller fellowship; B.K. by a T32 fellowship; O.E. by NIH UL1TR002384, R01CA194547, LLS SCOR 180078-02, 7021-20 and Janssen and Eli Lilly research grants; J.S. by HL119215; and Q.J.Z. by NIH R01DK106253, UC4DK116280, and the Neuroendocrine Tumor Research Foundation grant NTRF192269; J.R.S. by NIH R01HL119215. We thank R. Chavez for technical assistance and animal maintenance, the Visual Function Core at Weill Cornell Medicine for live-imaging resources, J. Shieh for HUVEC isolations, C. Wasserstein, M. Mishaan, M. Bell, C. Cheung and W. Gu for support with organoid cultures and A. Gjinovci for surgical help with the isolation of mouse small bowels.

Author contributions B.P. and S.R. conceived the study and wrote the manuscript. B.P., D.-H.T.N., R.S., K.S., R.E.S., S.Y.R. and S.R. discussed and analysed data. R.S. provided microscopy expertise. B.P., D.-H.T.N., G.L., R.S., Y. Liu, F.G., Y. Lin, J.M.G.-S., M.Y., S.Y.R. and S.R. performed experiments and analysed data. A.F.P., A.M.T., F.S. and P.D.C. carried out and analysed experiments on the decellularized intestines. Y. Liu, D.R., P.Z., T.Z., B.K., O.E. and J.Z.X. analysed ChIP–seq, RNA-seq and single-cell sequencing. M.W., T.H., S.L., L.D., J.S. and Q.J.Z. assisted with organoid cultures. All the authors read and provided feedback on the figures and manuscript.

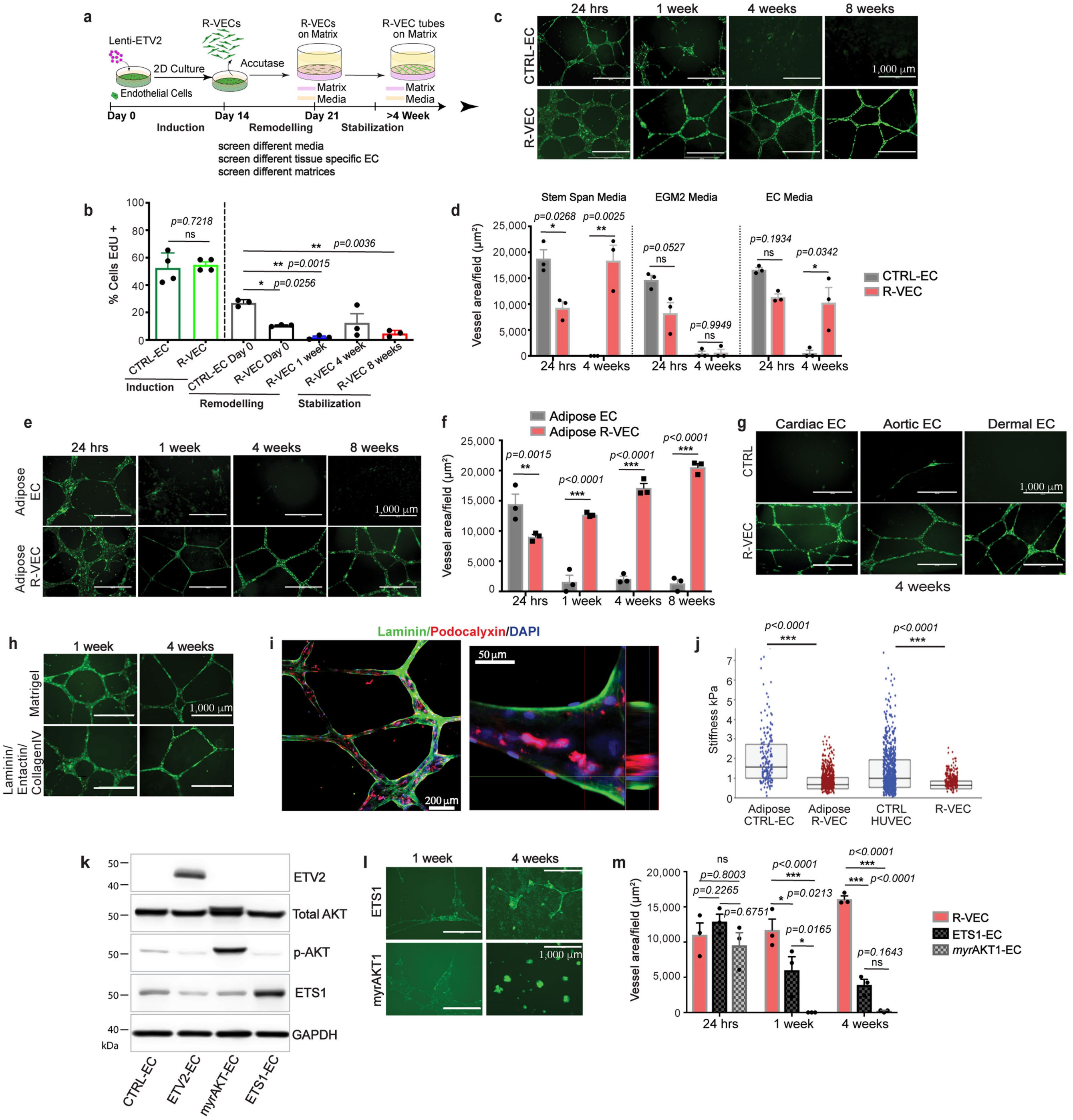
Competing interests S.R. is the founder and a non-paid consultant to Angiocrine Bioscience. O.E. is a scientific advisor and equity holder in Freenome, Owkin, Volastra Therapeutics and One Three Biotech. R.E.S. is a scientific advisor and member of the SAB for Miromatrix Inc.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2712-z.

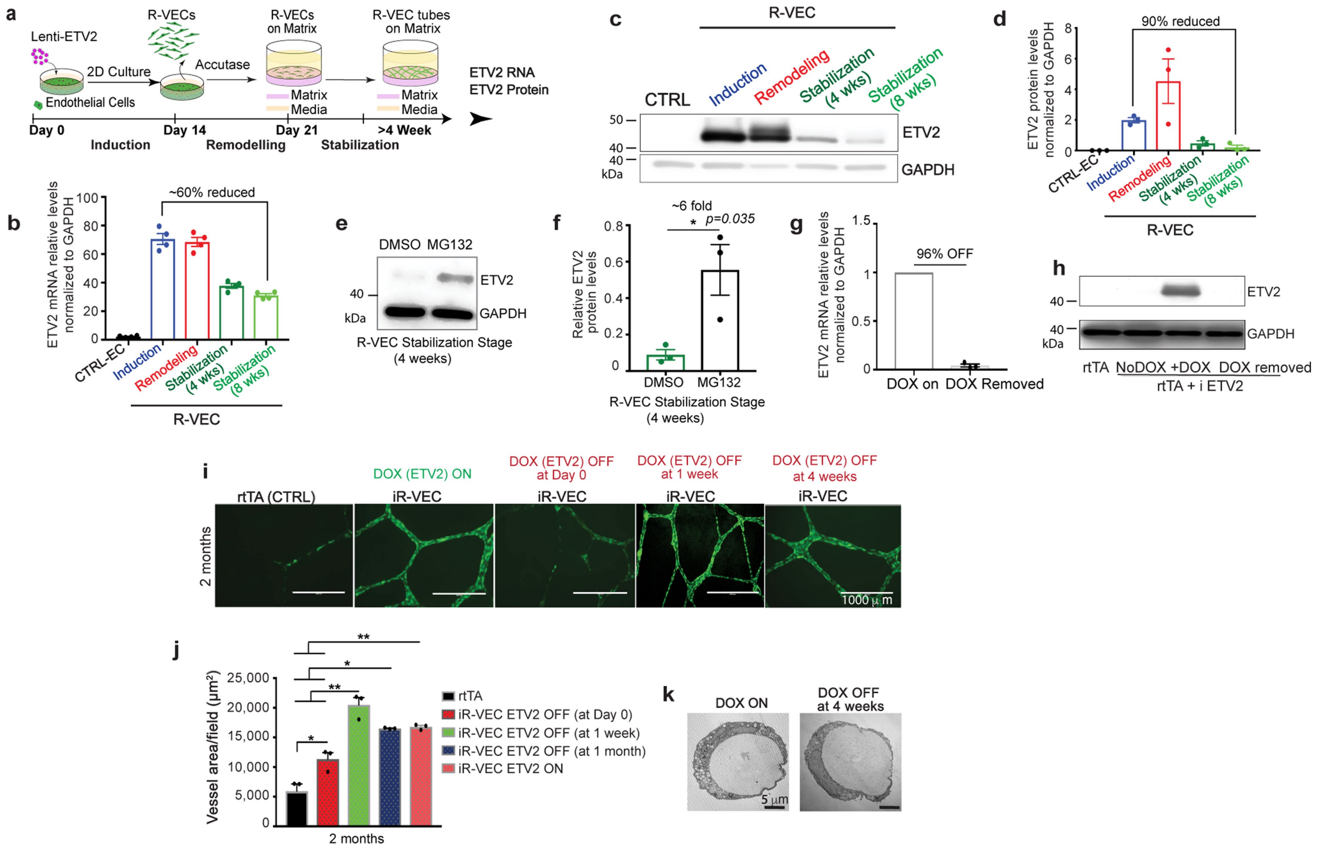
Correspondence and requests for materials should be addressed to S.R.

Peer review information *Nature* thanks Hans Clevers, Harald Ott and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

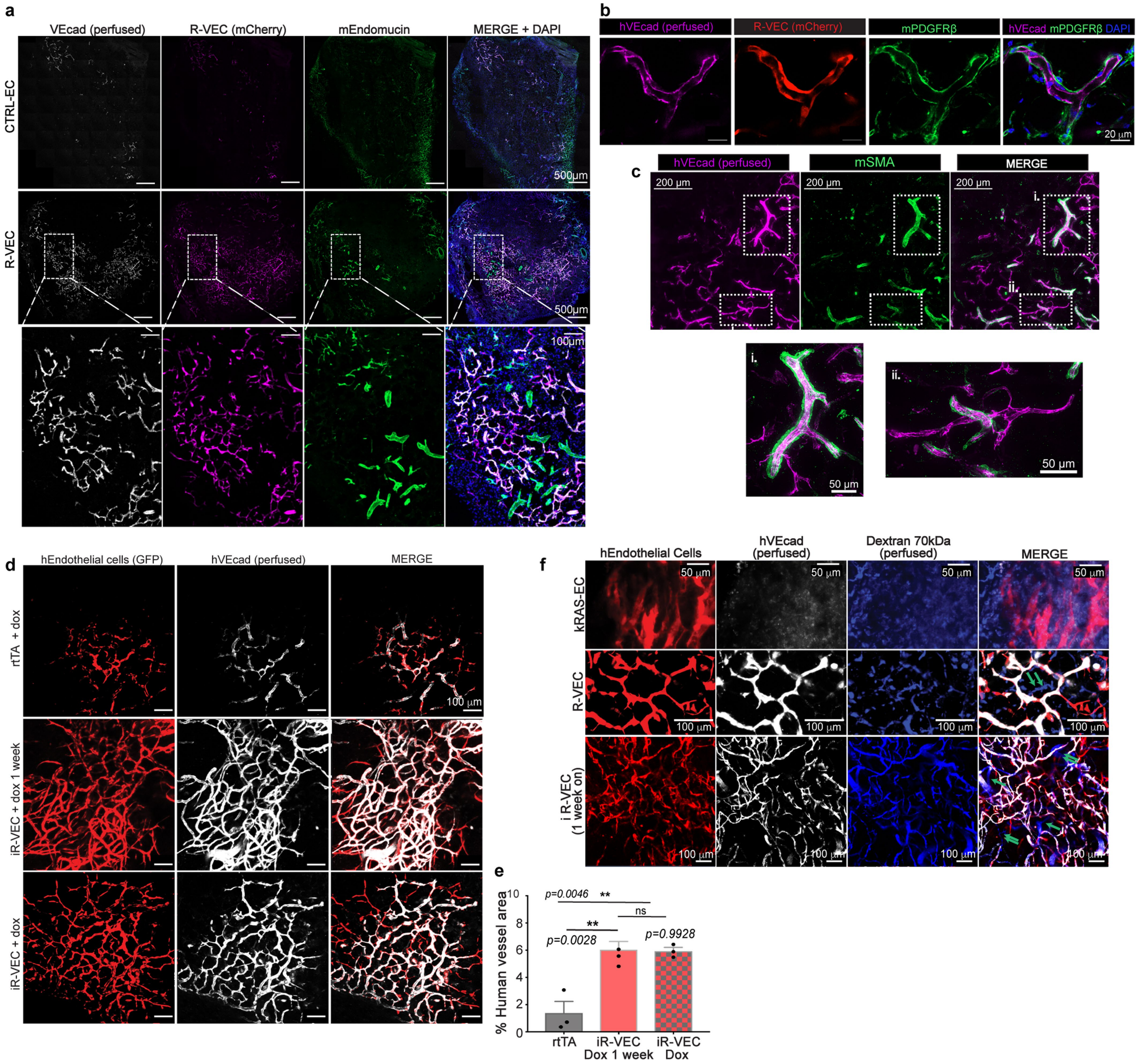


Extended Data Fig. 1 | See next page for caption.

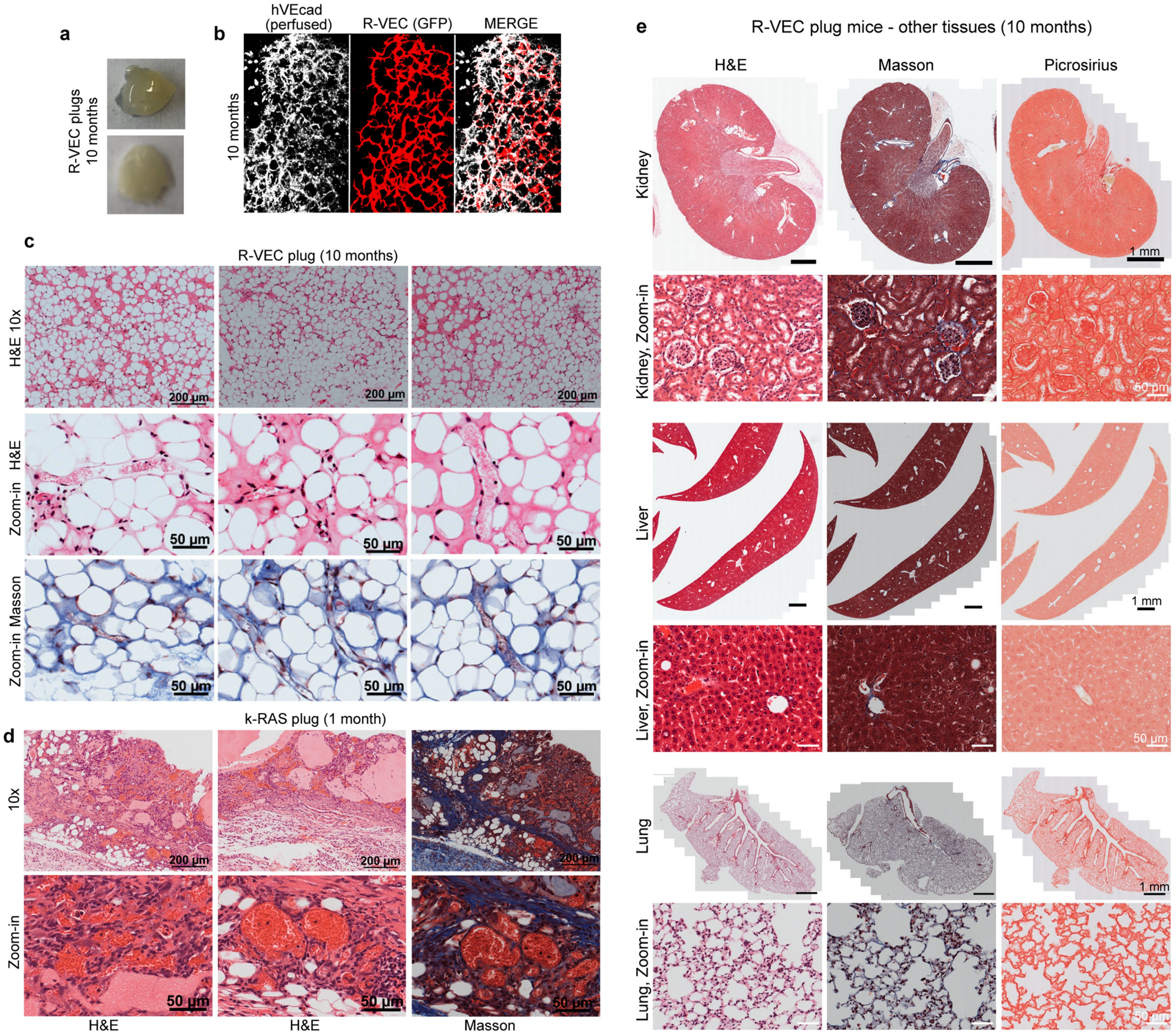
Extended Data Fig. 1 | ETV2 confers mature human ECs with the ability to autonomously self-assemble into lumenized, durable, branching and patterned vessels in vitro without the constraints of bioprinted scaffolds. **a**, Overview of experimental set-up for vessel formation in vitro for screen of different media, extracellular matrix components and tissue-specific ECs. **b**, The proliferation of GFP-transduced R-VECs and control ECs (CTRL-ECs) at each stage of vessel formation. EdU^+ cells were quantified after a 16-hour EdU pulse. **c**, Time course of vessel formation on Matrigel for GFP^+ CTRL-EC and R-VECs over 8 weeks. **d**, Vessel formation using R-VEC or CTRL-EC in three different enriched pro-angiogenic media (Supplementary Data 2): Serum-free StemSpan with Knockout serum replacement and Cytokines, EGM-2 and complete EC media on Matrigel. R-VEC formed the most robust lumenized vessels in serum-free StemSpan with knockout serum replacement medium and cytokines, as compared to other media with serum. CTRL-EC failed to form durable stable vessels. **e**, **f**, Time course (**e**) and quantification (**f**) of tube formation for GFP^+ human Adipose CTRL-EC and human Adipose R-VEC on Matrigel. **g**, Representative images of tissue-specific GFP^+ R-VEC and CTRL-EC isolated from adult human heart (cardiac EC), aorta (aortic EC) and skin (dermal EC) demonstrated robust and stable vessels at 4 weeks on Matrigel. **h**, Representative images of GFP^+ R-VEC vessels formed on Matrigel or a pre-defined matrix of laminin/entactin and collagen IV (LEC). **i**, Immunostaining of R-VEC-tubes displayed apicobasal proper polarity with podocalyxin, apical (in red) and laminin, basal (in green). The right image is an orthogonal projection. **j**, Stiffness measurements by atomic force microscopy (AFM) of adult Adipose and HUVEC ECs with and without ETV2. In both cases, ETV2-transduced ECs are significantly less stiff than their counterparts. The abbreviated box plots indicate the interquartile range and median for each condition. **k**, HUVECs were transduced with either an empty lentiviral vector or lentiviral vectors with ETV2, myrAKT or ETS1 constructs, and used in a vessel formation assay. Western Blot analysis for expression of ETV2, p-AKT, total AKT and ETS1 in those cells. **l**, Representative images for ETS1 or myrAKT1 transduced GFP^+ HUVECs in a vessel formation assay on Matrigel. **m**, Quantification of vessel area for ETS1, myrAKT1 and ETV2 (R-VEC) cells indicated that ETS1-EC and myrAKT1-EC fail to form robust vessel formation as compared to R-VEC. Data are mean ± s.e.m. NS, not significant; $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. For statistics, see Supplementary Data 1. For medium formulations, see Supplementary Data 2.



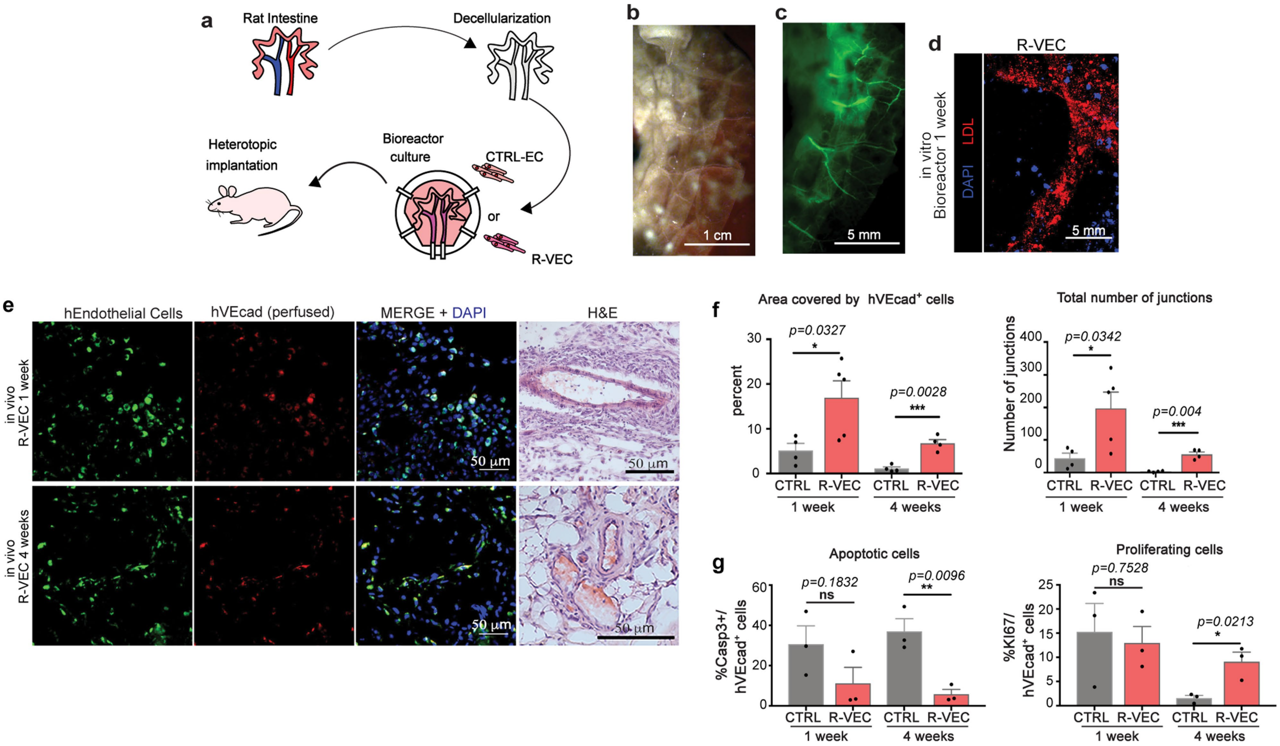
Extended Data Fig. 2 | Transient ETV2 expression in adult human ECs is sufficient for the generation and maintenance of durable long-lasting R-VEC vessels in vitro. a, Schematic for ETV2 mRNA and protein levels assessment at each of the three stages of R-VEC vessel formation. **b**, Quantification of ETV2 mRNA levels at each stage of vessel formation. **c**, **d**, Western blot analysis (**c**) and densitometric quantification (**d**) of ETV2 protein levels at each stage of vessel formation. GAPDH was used as a loading control. **e**, A proteasome inhibitor (MG132) restored ETV2 levels by ~sixfold when added to R-VECs during the stabilization stage. **f**, Densitometric quantification of western blots in **e**. **g**, qRT–PCR (**g**) and western blot (**h**) assessment of ETV2 levels after doxycycline removal. **i**, Representative images of GFP^{+} iR-VECs on Matrigel with inducible ETV2 expression at 2 months. ETV2 was turned off at day 0, day 7 and at 4-weeks post start of the remodelling stage 2. **j**, Quantification of iR-VEC vessels at 2 months. **k**, Electron microscopy pictures of a lumen present both in vessels in which doxycycline was continuously present and in vessels in which doxycycline was removed after 1 month. Data are mean ± s.e.m. NS, not significant; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. For statistics, see Supplementary Data 1. For medium formulations, see Supplementary Data 2.



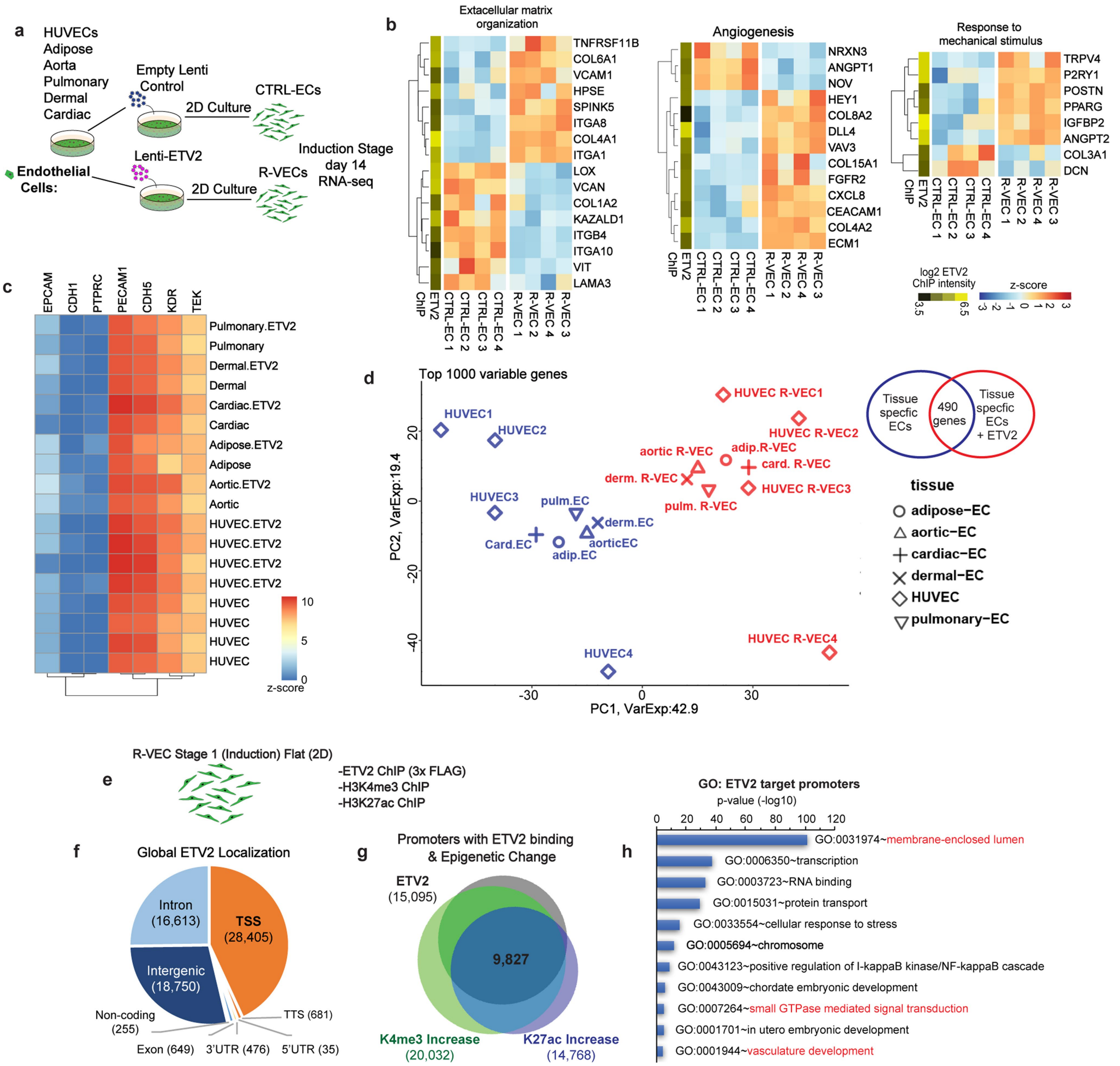
Extended Data Fig. 3 | R-VEC vessels are functionally anastomosed to host vessels and not leaky in vivo. **a**, Fluorescently labelled R-VEC or CTRL-EC cells in LEC matrix were subcutaneously injected in the flank of SCID beige mice and retrieved at 2 months. Human-specific VEcad antibody (hVEcad) was injected intravitally right before euthanasia. Sections of the plugs were stained for mouse ECs with an anti-mouse endomucin antibody (mEndomucin), identifying properly organized human R-VECs anastomosing with mouse vessels (thickness = 50 µm). Sections were also stained with the nuclear stain DAPI. **b**, **c**, Plugs in **a** were post-stained with hVEcad and a mouse PDGFRβ antibody (**b**) or mouse SMA antibody (**c**) (thickness = 50 µm). **d**, In vivo plug assay, in which mice were subcutaneously injected with either control ECs (HUVECs transduced only with rtTA lentivirus) or stage 1 doxycycline-inducible-ETV2 ECs (iR-VECs: HUVECs transduced with both rtTA and inducible ETV2 lentivirus) in LEC matrix. One group of mice was on doxycycline (ETV2 continuously on) and another group of mice was on doxycycline food diet for 1 week (ETV2 on) and then switched to regular food (ETV2 off). All mice were euthanized 2 months post-implantation. Red indicates the GFP labelled human ECs, white: Anti-VEcad antibody that was retro-orbitally injected before euthanizing the mice. **e**, Quantification of vessel area for rtTA only plugs, mice on doxycycline for 1 week, and mice continuously on doxycycline diet (ETV2 on). All mice were euthanized 2 months post-implantation. **f**, 70 kDa fluorescent dextran (in blue) and human VEcad (in white) were injected in mice implanted with fluorescently labelled R-VECs (in red, 5-months post-implantation), iR-VECs (in red, 1 week on doxycycline food and euthanized at 2 months) or K-RAS-HUVECs (K-RAS-EC) (in red, 2-weeks post-implantation) to assess anastomosis and leakiness of vessels. K-RAS-EC vessels showed dextran leakiness, whereas R-VEC and iR-VEC vessels exhibited patency and non-leakiness. Green arrows point at perfused mouse vessels that were also perfused with dextran. Data are mean ± s.e.m. NS, not significant; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. For statistics, see Supplementary Data 1. For medium formulations, see Supplementary Data 2.



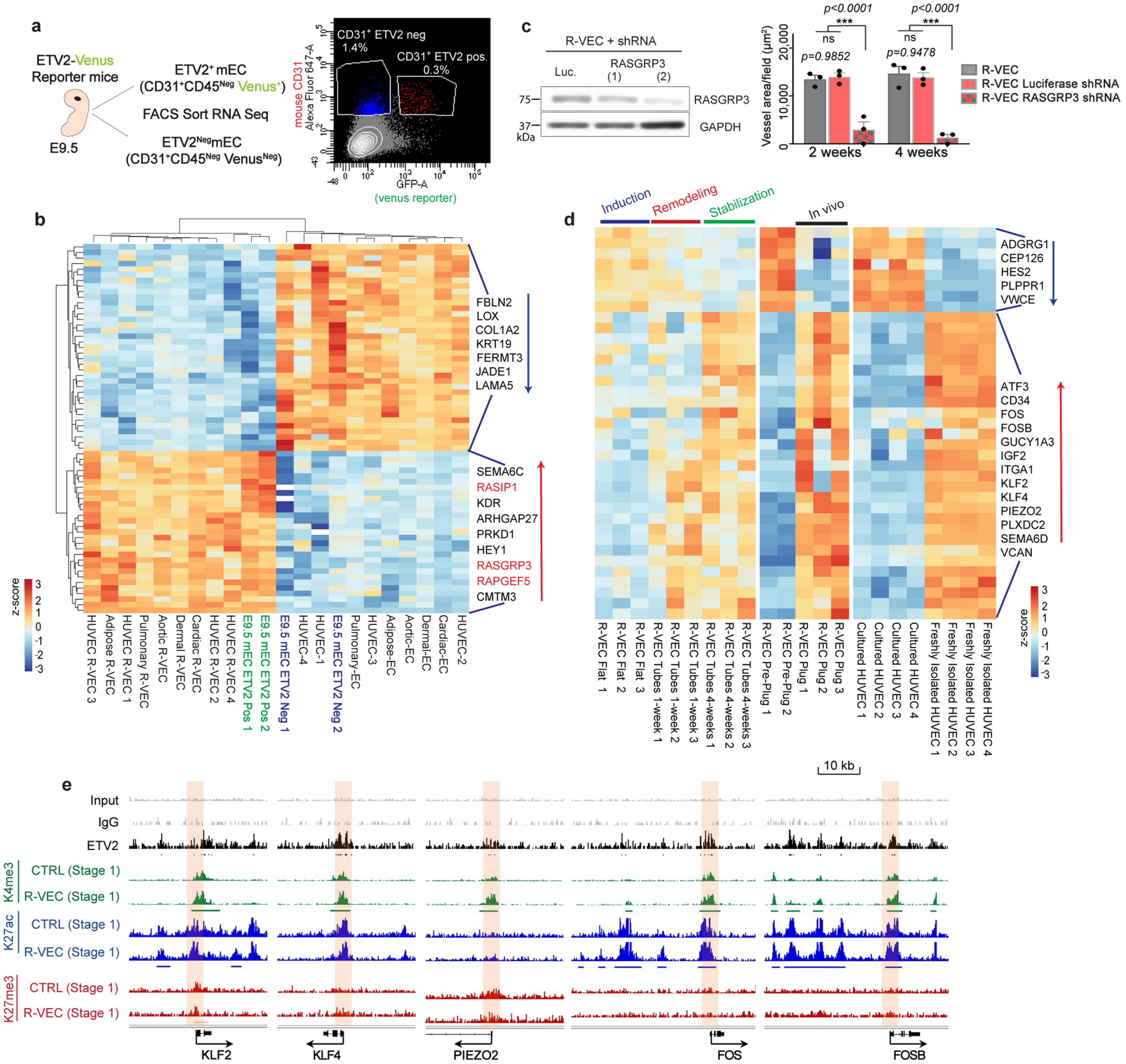
Extended Data Fig. 4 | Implanted R-VECs form stable, patterned, branching and durable vessels in vivo without features of vascular malformations, cysts, adenomas, haemangiomas or metastasis. **a**, Representative images of non-haemorrhagic R-VEC plugs at 10 months. **b**, Whole-mount microscopy of R-VEC plugs at 10-months post perfusion with anti-human VEcad antibody (hVEcad). **c**, **d**, Representative H&E and Masson staining of R-VEC plugs at 10 months (**c**). There were no features of cysts or haemangiomas present, in contrast to KRAS-EC plugs (at 4 weeks) that formed an EC tumour (**d**). **e**, There was no metastasis of R-VECs to other tissues 10 months after plug implantation and the tissues were assessed to be normal without fibrosis and architectural disruption or tumorigenesis as evaluated by H&E, Masson and picrosirius staining.



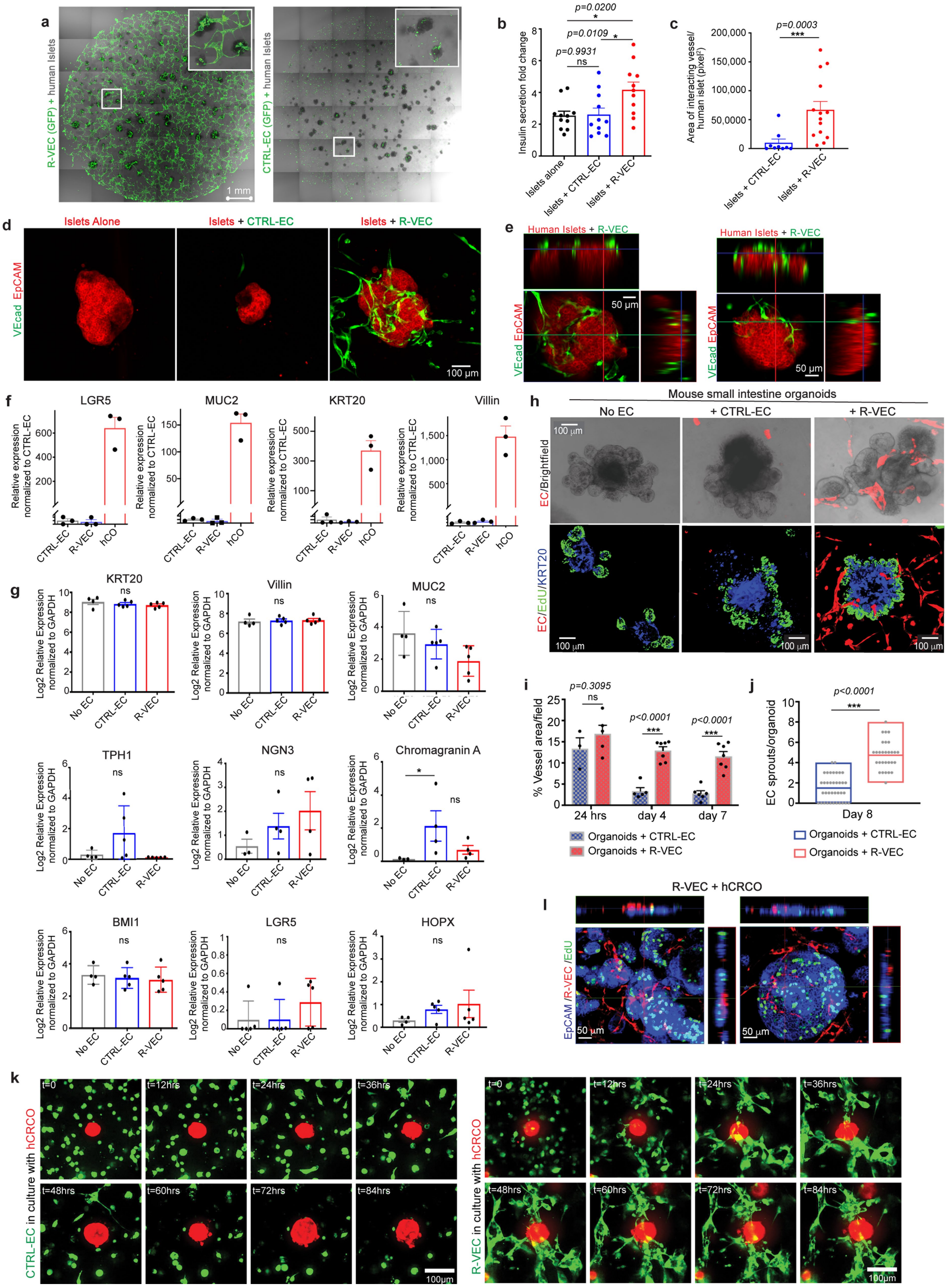
Extended Data Fig. 5 | Decellularized intestinal scaffolds re-endothelialized with R-VECs engraft in vivo after omental implantation. a, Schematic of experimental procedure for heterotopic implantation of decellularized intestinal scaffold vascularized using R-VECs. **b**, Rat intestines were cannulated through lumen, mesenteric artery and mesenteric vein. **c**, Decellularized intestine preserves native vasculature (green = GFP$^+$ R-VECs). **d**, Seeded GFP labelled R-VECs spread evenly and reach distal capillaries. **e**, Heterotopic implantation of re-endothelialized intestines in immunodeficient mice omentum shows engraftment after 1 and 4 weeks of GFP$^+$ R-VECs and anastomosis to the host vasculature as indicated by intravital intravenous injection of anti-human VEcad antibody (hVEcad). Representative H&E stainings show anatomical normal perfused vessels. **f**, Quantification of the area covered by R-VEC compared to CTRL-EC in implanted re-endothelialized intestines at 1 week and 4 weeks. **g**, Quantification of R-VEC and CTRL-EC proliferation and apoptosis in implanted re-endothelialized intestines at 1 and 4 weeks. Data are mean ± s.e.m. NS, not significant; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. For statistics, see Supplementary Data 1.



Extended Data Fig. 6 | ETV2, by directly binding to promoters and enhancers of target genes, regulates differentially expressed genes in R-VECs. a, Schematic of RNA-sequencing performed on R-VECs and CTRL-ECs derived from different tissue-specific ECs during stage 1 induction phase (2D monolayers). **b**, R-VECs or CRTL-ECs were analysed by RNA sequencing. Heat maps of selected genes within top enriched GO categories. Values are $\log_2$-normalized CPM, centred and scaled by row. ETV2 binding from ChIP–seq at the promoter of each differentially expressed gene is shown in the yellow-green heat map. **c**, R-VECs retain essential EC fate genes at stage 1 induction phase across all tissue-specific ECs. The data are presented as $\log_2$(CPM) with no scaling by row or column. **d**, PCA plot based on the top 1,000 most variable genes across ECs with and without ETV2 from different tissues during stage 1 induction, using $\log_2$-normalized CPM after subtracting tissue-specific effects using limma's removeBatchEffect function. **e**, ETV2 ChIP was performed on R-VECs using an anti-flag antibody at the induction stage 1 (2D), along with histone modification ChIP for H3K4me3 and H3K27ac. Enriched regions were analysed by ChIP–seq. **f**, Genomic distribution of ETV2 peaks in R-VEC (Stage 1). The number in brackets is the number of ETV2 peaks in each region. **g**, Promoters bound by ETV2 have an increase in both K4me3 and K27ac. **h**, GO enrichment in genes with ETV2 binding at promoters. For statistics, see Supplementary Data 1.

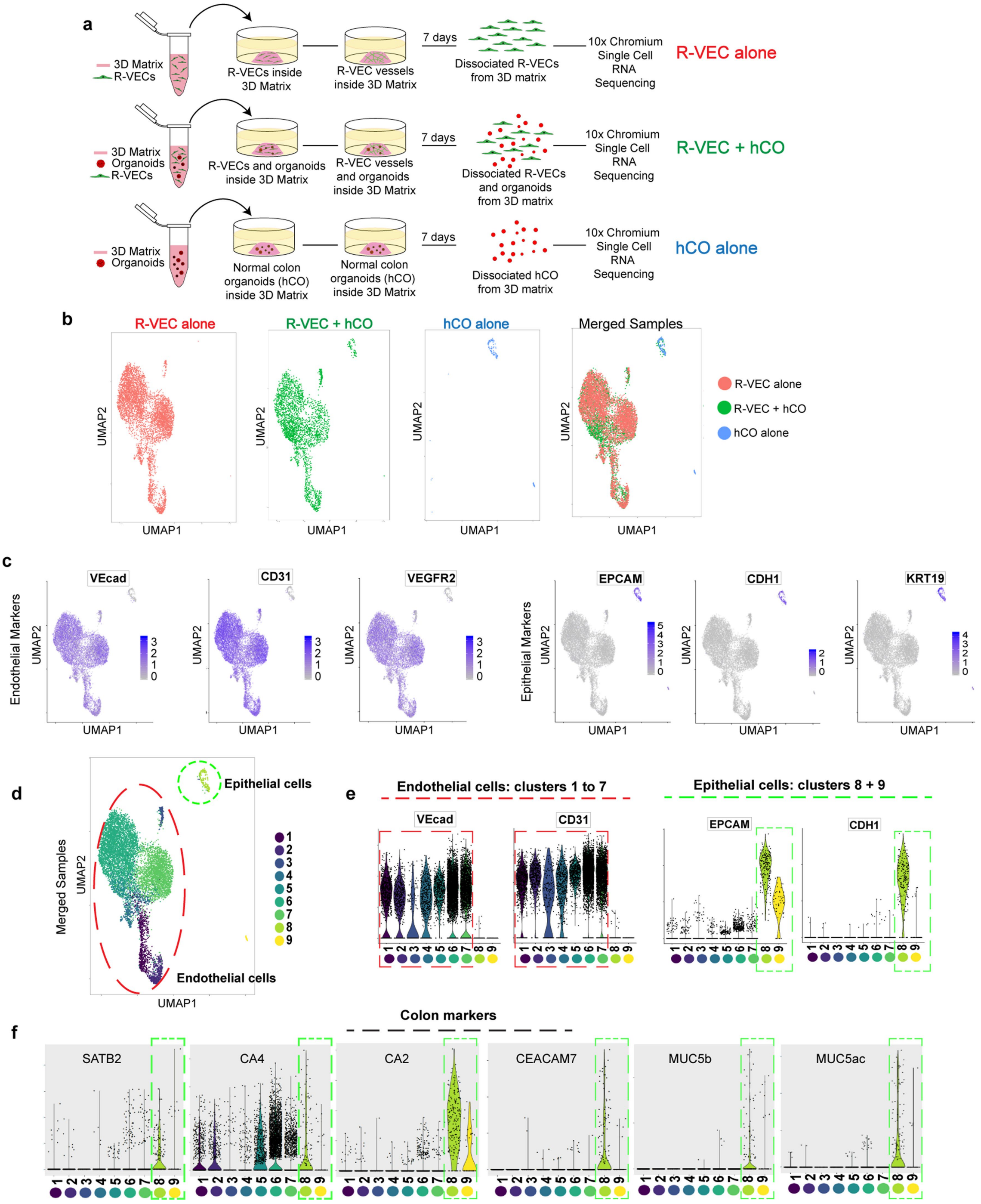


Extended Data Fig. 7 | ETV2 in R-VECs endows ECs with transcriptional adaptability and plasticity. **a**, Diagram of EC sample preparation from ETV2 Venus reporter mice by FACS sorting. ETV2$^+$ and ETV2$^-$ ECs were sorted at day E9.5. ECs were sorted as non-haematopoietic CD31$^+$CD45^{neg} cells. **b**, Heat map of overlap of differentially expressed genes in ETV2$^+$ vs. ETV2$^-$ ECs at E9.5 and R-VECs (stage 1) vs. CTRL-EC from different tissues, using tissue-adjusted log$_2$(CPM), centred and scaled by row. **c**, Knockdown of RASGRP3 by two different shRNAs in R-VECs, shRNA against Luciferase was used as control. Vessel quantification upon *RASGRP3* knockdown. **d**, Heat map displaying overlapping differentially expressed genes from R-VEC at stabilization stage 3 (4 weeks) vs. R-VEC at induction stage 1, R-VECs in vitro pre-plug (stage 1 induction stage) vs. R-VECs in vivo in plugs (1 month), and freshly isolated vs. cultured HUVECs. Values represent tissue-adjusted log$_2$(CPM), centred and scaled by row. **e**, ChIP–seq depicting genes that are differentially expressed in the stabilization stage 3 phase, but that are already directly bound by ETV2 and epigenetically primed for expression at induction stage 1 (2D monolayers). ETV2 ChIP–sequencing was performed on R-VECs using an anti-flag antibody. Mouse IgG was used as a control for ETV2 ChIP. Histone modification ChIP for H3K4me3, H3K27ac and H3K27me3 was performed on both CTRL-EC and R-VEC at the induction stage 1 (2D monolayers). Enriched regions were analysed by ChIP–sequencing. Black bar, ETV2 enriched regions. Green bar, the region with increased K4me3 modification. Blue bar, the region with increased K27ac modification. Promoter regions bound by ETV2 are highlighted in cream. Track range ETV2/K27me3/K27ac/, 0–0.3; K4me3/input/IgG, 0–1. For statistics, see Supplementary Data 1. For medium formulations, see Supplementary Data 2.

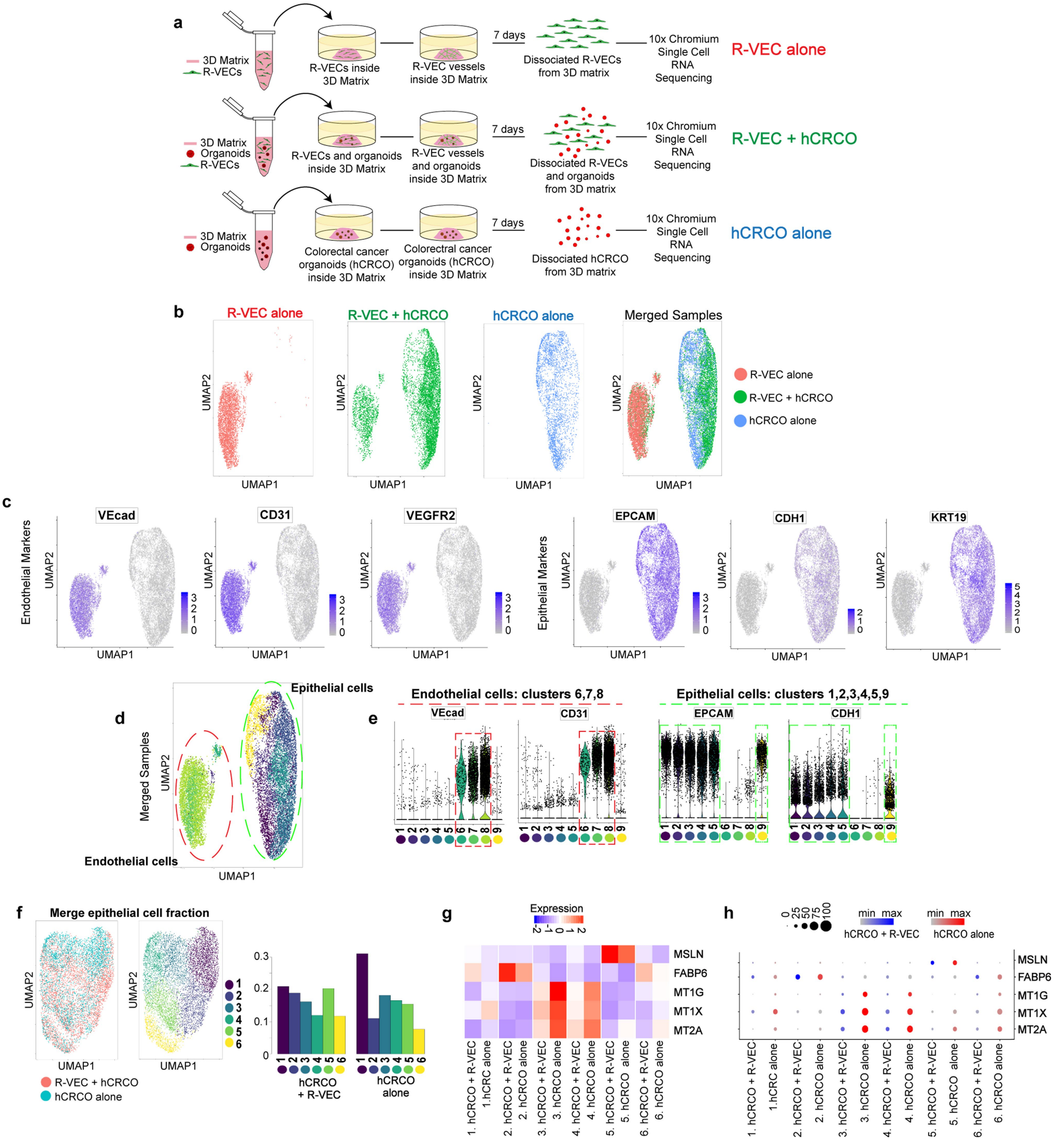


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | R-VECs physiologically arborize human pancreatic islet explants and organoids. a, Human islet explants were cultured in Matrigel droplets (volume 50 µl) either with GFP labelled CTRL-EC or R-VEC (day 4). **b**, Insulin secretion fold change after glucose stimulation at 16.7 mM vs. 2mM glucose (2-week time point). **c**, Vessel area of ECs directly interacting with islets at week 2. **d**, EpCAM and VEcad staining of islets co-cultured in a Matrigel droplets at 2 weeks. **e**, Orthogonal projections of R-VECs in co-culture with human islets at two weeks, demonstrating strong interaction of the sprouting R-VEC vessels with islets. **f**, Human COs were derived from isolated crypts from colon biopsies of healthy human donors. Colon organoids were confirmed to express proper markers by quantitative RT–PCR. **g**, Quantitative RT–PCR of various colon markers for human COs, co-cultured with CTRL-EC or co-cultured with R-VEC for 8 days. Epithelial cells were sorted out as live CD31^{neg} non-vascular cells. **h**, Mouse small intestine organoids were cultured alone, or in the presence of CTRL-EC or R-VEC (day 8). Confocal representative images of EdU^{+} (proliferating cells), KRT20^{+} (differentiated epithelial cells in blue) and ECs (mCherry - red) of co-culture experiment with mouse intestinal organoids. **i**, Quantification of vessel area over the course of 7 days in co-cultures of mouse intestine organoids with CTRL-EC or R-VEC. **j**, Vessel arborization quantified as EC sprouts in direct contact/organoid in CTRL-EC versus R-VEC wells. **k**, Time-lapse representative images show the progression of interacting ECs with CRCOs. CTRL-EC (in green) did not interact with CRCOs (in red) (top panel), whereas R-VEC (in green) form robust EC tubes to tap and wrap CRCOs (in red) (bottom panel). **l**, Orthogonal projections of CRCOs co-cultured with R-VECs (day 8). Data are mean ± s.e.m. NS, not significant; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. For statistics, see Supplementary Data 1. For medium formulations, see Supplementary Data 2.



Extended Data Fig. 9 | Endothelial and epithelial cell identification by scRNA-seq from co-cultures of normal COs with R-VECs. a, Schematic of 10X Chromium scRNA-seq experiments of R-VECs alone, R-VECs co-cultured with human COs, or COs alone. Samples were analysed 7 days post co-culture. The same compatible medium was used across all three conditions. **b**, UMAP of cells from each condition alone and the three conditions merged. **c**, Endothelial cells were identified as cells expressing either *VEcad*, *CD31* or *VEGFR2* and negative for the epithelial marker *EPCAM*. Epithelial cells were defined as positive for *EPCAM* and negative for any of the EC markers *VEcad*, *CD31* or *VEGFR2*. **d**, UMAP of the 9 unique clusters identified in the merged samples. **e**, Endothelial and epithelial cell specific markers were used to confirm the EC clusters (clusters 1 to 7) vs. epithelial cell clusters (clusters 8 and 9). **f**, The identity of epithelial cells in clusters 8 and 9 was confirmed as colon-specific by expression of marker genes including *SATB2*, *CA4*, *CA2* and others. For statistics, see Supplementary Data 1. For medium formulations, see Supplementary Data 2.



Extended Data Fig. 10 | Endothelial and epithelial cell identification by scRNA-seq from co-cultures of CRCOs with R-VECs. **a**, Schematic of 10X Chromium scRNA-seq experiments of R-VECs alone, R-VECs co-cultured with human CRCOs or CRCOs alone. Samples were analysed 7 days after co-culture. The same compatible medium was used across all three conditions. **b**, UMAP of cells from each condition alone and the three conditions merged. **c**, Endothelial cells were identified as cells expressing either *VEcad*, *CD31* or *VEGFR2* and negative for the epithelial marker *EPCAM*. Epithelial cells were defined as positive for *EPCAM* and negative for any EC markers *VEcad*, *CD31* or *VEGFR2*. **d**, UMAP of the 9 unique clusters identified in the merged samples. **e**, Endothelial and epithelial cell-specific markers were used to confirm the endothelial cell clusters (clusters 6, 7, 8) vs. epithelial cell clusters (clusters 1, 2, 3, 4, 5, 9). **f**, UMAP of merged epithelial cell fractions from hCRCO cultured alone or co-cultured with R-VECs. Six unique clusters were identified. **g**, **h**, Heat map (**g**) and dot plot (**h**) of differentially expressed genes in tumour epithelial cells in cluster 2 and cluster 5 that are enriched in co-culture with R-VECs. Differential expression was performed using the Wilcoxon rank-sum test; FDR-adjusted $P < 0.05$. For statistics, see Supplementary Data 1. For medium formulations, see Supplementary Data 2.

nature research

Corresponding author(s): Shahin Rafii

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Zen Black 2012 and Zen Blue 2 and 2.5 were utilized for image collection. BD FACS Diva V8.0.1 was utilized for FACS data collection.

Data analysis	GraphPad Prism 7.0 as used for all statistical analyses. Image J (Fiji Version 1.0) was used for image analysis/calculations. Zen Black 2012 and Zen Blue 2 and 2.5 were utilized for image processing. AngioTool v 0.6a IGV2.3.94 ,DAVID 6.7, HOMER 4.10.4 were utilized for ChiP analysis. MATLAB 2018 was used to analyze the interaction between organoids and endothelial cells STAR v2.6.0c for alignment of bulk RNA-seq data. featureCounts v1.6.2 for counting the number of reads mapping to each gene in bulk RNA-seq data. RSEM v1.2.28 for quantification of isoform level counts for bulk RNA-seq data. FastQC v0.11.7 for QC of bulk RNA-seq data. QoRTs v1.3.0 for QC of aligned, bulk RNA-seq data. DESeq2 v1.18.1 for differential gene expression analysis of bulk RNA-seq data. limma v3.34.9 for removal of batch effects prior to some visualizations related to bulk RNA-seq . Seurat V2.3.4 for single cell analysis R 2.3.4 and 3.5.2 were used for Bulk and Single cell RNA-sequencing, ChIP-sequencing

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The raw data for Figures 1-4 and Extended Figures 1,2,3,5,7,8 are provided with the paper. The RNA-sequencing data can be viewed on GEO under the record GSE131039. The ChIP-sequencing data can be viewed on GEO under the record GSE147746. The single cell RNA-sequencing data can be viewed on GEO under the record GSE148996.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were utilized to determine sample size. Sample size was determined based on previous experience in the lab and previous publications. All experiments were repeated independently 3 times, unless otherwise noted. Sample size for each experiment is included in Supplementary Data 1.
Data exclusions	There was no data exclusion.
Replication	Attempts at replication have been successful. We have tested our system across more than n=10 R-VEC lines with several virus preparations, across several years, and lot numbers for commercially available materials.
Randomization	Samples and animals were allocated randomly in each experiment.
Blinding	For zonal confocal miscroscopy, the investigator setting up the time-lapse, picking the organoids to be imaged was blinded. This was done to ensure that there was no bias in the organoids imaged based on size or endothelial cells around organoids at time 0. For other experiments no blinding was done. In part, blinding was difficult in most experiments due to the obvious differences in vessel formation between CTRL-EC and R-VEC. Indeed, R-VECs in most experiments manifested remarkable capacity to establish lumenized vascular network, rendering blinding of experiments impractical. Experiments in the paper were quantified utilizing standardized quantitative methods to avoid bias.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	Normal and tumor organoids are unique to the patients/human subjects they were isolated from. We have been able to repeat our experiments across different organoid lines. Most of our tumor organoids were procured from the stocks at the Institution from Precision Medicine at Weill Cornell Medicine.

Antibodies

Antibodies used	Anti-human VE-cadherin (BV9 clone) Biolegend 348514 Retro-orbital injection (25µg suspended in 100µL 1xPBS/mouse) Anti-human CD31 Biolegend 303124 Flow cytometry: 10µg/ml Anti-ETV2 Abcam Ab181847 WB: 1:1000 Anti-ETS1 Abcam Ab225868 WB: 1:1000 Anti-mouse PDGFRβ Biolegend 136004 IF: 1:500 Anti-mouse SMA Abcam ab5694 IF: 1:200 Anti-mouse Endomucin Santa Cruz sc-65495 IF: 1:100 Anti-human CD31 (clone WM59) BD Biosciences 561654 IF: 1:100 Anti-RASGRP3 Cell Signaling 3334S WB: 1:1000 Anti-Keratin 20 Cell Signaling 13063S IF: 1:200 Anti-EpCAM Biolegend 324212 IF: 1:100 Anti-AKT Cell Signaling 4691S WB: 1:5000 Anti-phospho-AKT Cell Signaling 4060S WB: 1:2000 Anti-GAPDH Cell Signaling 5174S WB: 1:10000 Anti-mouse CD31 (Clone 390) Biolegend 102418 Flow Cytometry (1µl/1million cells) Anti-mouse CD45 (clone F30-11) Biolegend 103124 Flow Cytometry (1µl/1million cells) Anti-Ki67 Abcam AB15580 IF: 1:200 Anti-Cleaved Caspase3 Cell Signaling 9661S IF: 1:100 Mouse Isolectin B4 ThermoFisher 132450 Retro-orbital injection, 50µl/mouse Dextran (70 kDa) ThermoFisher D1818 Leakiness test, 1.25 mg/mouse in 125 µl PBS Anti-H3K4me3 Abcam Ab8580 ChIP: 7.5µg for 1x107 cells Anti-H3K27ac Abcam Ab4729 ChIP: 7.5µg for 1x107 cells Anti-Flag Sigma F1804 ChIP: 7.5µg for 1x107 cells Anti-H3K27me3 Millipore 07-449 ChIP: 1 µg for 10,000 cells Anti-Insulin Abcam ab7842 IF: 1:100 Anti-VEcad R&D AF938 IF: 1:100
Validation	The following antibodies were validated in our experiments: 1) For Fig. 1f-h, Extended Fig. 3a-f, Extended Fig. 4b and Extended Fig. 5e, the injected BV9 (Biolegend) anti VEcad antibody was validated as specifically staining the human endothelial cells in the plugs/transplanted intestines, as the antibody staining specifically matched the fluorescent (GFP or mCherry) of the injected human endothelial cells . 2)For Extended Fig. 2c,e,h ETV2 antibody for Western Blots (Abcam) was validated by overexpressing an ETV2-flag construct. Other antibodies were not validated. Manufacturer's guidelines about concentrations and expected results were followed.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T (ATCC) R-spondin1 overexpressing line (derived in the laboratory of Calvin Kuo) L-WRN overexpressing cell line (derived in the laboratory of Thaddeus Stappenbeck)
Authentication	No authentication was performed.
Mycoplasma contamination	CTRL-HUVECs and ETV2-HUVECs were routinely checked for mycoplasma and were found to be negative. Other cells/cell llines were not tested for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register)	Cells used in this study are not among the commonly misidentified lines.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Experiments utilizing the following animals were performed under the approval of Weill Cornell Medicine Institutional Animal Care and Use Committee (IACUC), New York, NY. -SCID Beige mice (male and female, 8-12 weeks old) from Taconic were used for implants and ischemic limb experiments. -ETV2-Venus reporter mice were a kind gift of Dr. Valerie Kouskoff. Embryos from E9.5 pregnant female ETV2/+ reporter mice were used to isolate endothelial cells with or without ETV2 expression. Only embryos positive for ETV2 reporter were used. Animals used for decellularization experiments were maintained and experiments performed in accordance with the UK Animals (Scientific Procedures) Act 1986 and approved by the University College London Biological Services Ethical Review Process (PPL 70/7622). Animal husbandry at UCL Biological Services was in accordance with the UK Home Office Certificate of Designation. -NOD-SCID-gamma mice (male and female, 12 weeks old), bred at University of College London, were used for transplantation of decellularized intestines. -Sprague Dawley rats (male and female, 6-12 months, 250-350 g) were utilized for harvesting intestines for the decellularization experiments.
Wild animals	The study did not involve wild animals.
Field-collected samples	This study did not involve field-collected samples.

Human research participants

Policy information about studies involving human research participants

Population characteristics	HUVECs were isolated from human umbilical cords obtained as left over discarded tissues at the New York Presbytarian Hospital. The population are healthy full term pregnant women who have either gone Caesarian section or normal delivery. Fat endothelial cells were isolated from human fat tissue obtained from leftover tissue after reconstruction surgery at the New York Presbytarian Hospital. The population is healthy adult individuals. Normal and adenoma tissues were collected from colonic resections.
Recruitment	The IRB at Weill Cornell Medicine deemed the studies on HUVECs exempt from the requirement of informed consent. As umbilical cords are deemed discarded tissues, the recruitment does not require informed consent and is obtained through the hospital personnel depending on the availability of the discarded and left over umbilical cords. Fat endothelial cells, normal and adenoma tissues from colonic resection were collected according to protocols approved by Weill Cornell Medicine Institutional Review Board following appropriate consent.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	The ChIP-sequencing data can be accessed on GEO under GSE147746.
Files in database submission	See Supplementary Table 1
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	See Supplementary Table 1
Sequencing depth	All ChIP-seq files are generated as single-end 51 bp reads. The information about sequencing depth in each ChIP-seq file was attached in Supplementary Table 1.
Antibodies	All antibodies used are attached in the "Antibodies" section of the reporting summary.

Peak calling parameters

Command line for ChIP-seq read alignment:

```
bwa aln -t 4 hg19bwaidx file.fastq.gz > file.bwa
bwa samse hg19bwaidx file.bwa file.fastq.gz > file.sam
samtools view -bS file.sam > file.bam
samtools sort file.bam -o file.sort
samtools index file.sort.bam
java -Xmx2g -Dsnappy.disable=true -jar picard-tools-1.69/MarkDuplicates.jar INPUT=file.sort.bam OUTPUT=file.sort.rd.bam
```

```
REMOVE_DUPLICATES=true METRICS_FILE=file.rd.txt AS=true VALIDATION_STRINGENCY=LENIENT
```

Command line to identify ETV2 peak by MACS2 (p-value<0.05):

```
macs2 callpeak -t ETV2_ChIP.bam -c ETV2_input.bam -f BAM -g hs --outdir ETV2_peak -p 0.05
```

Command line to identify K27me3 peak with differential sequence intensity in different cell types by SICER (W=200, G=600, FDR=0.01):

```
sh SICER-df.sh HUVEC_ETV2_tube_K27me3.bed HUVEC_ETV2_tube_input.bed HUVEC_ETV2_K27me3.bed
HUVEC_ETV2_input.bed 200 600 .01 .01
```

Data quality	MACS2 was used to identify genomic enrichment (peak) of ETV2 from the ChIP-seq data, with sequencing data from input DNA as control. We identified 24,570 ETV2 peaks in total with p-value < 0.05.
Software	IGV2.3.94 was utilized for ChIP analysis. BWA (version 0.5.9) was used for ChIP-seq reads alignment. MACS2 was used to identify genomic enrichment (peak) of ETV2 from the ChIP-seq data, with sequencing data from input DNA as control. SICER (version 1.1) was used to identify genomic enrichment (peak) with different K27me3 modification in different cell types, with sequencing data from input DNA as control. HOMER was used to comput the read counts in individual promoters.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Embryos from ETV2 reporter mice were prepared as described in the methods.
Instrument	BD FACS AriaII
Software	FACS Diva for collection/analysis
Cell population abundance	FACS sorting efficiency could not be performed by flow-cytometry due to low cell number. The cells were submitted for RNA-seq and ETV2 expression was found only in the ETV2 positive sorted fraction. CD31 as expected was found on both sorted populations.
Gating strategy	FSC-A/SSC-A for mononuclear cells followed by FSC-H/FSC-W and SSC-H/FSC-H for singlets, DAPI negative for live cells, - CD45 negative, then double positive for CD31[APC] and ETV2 reporter [venus] were sorted for ETV2 positive ECs or positive only for CD31[APC] but negative for ETV2 reporter [venus] were sorted for ETV2 negative ECs (Extended Figure 8c)

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Mechanics of a multilayer epithelium instruct tumour architecture and function

https://doi.org/10.1038/s41586-020-2695-9

Received: 4 September 2019

Accepted: 22 July 2020

Vincent F. Fiore[1], **Matej Krajnc**[2,3], **Felipe Garcia Quiroz**[1,7], **John Levorse**[1], **H. Amalia Pasolli**[4], **Stanislav Y. Shvartsman**[2,5,6] **& Elaine Fuchs**[1]✉

Loss of normal tissue architecture is a hallmark of oncogenic transformation[1]. In developing organisms, tissues architectures are sculpted by mechanical forces during morphogenesis[2]. However, the origins and consequences of tissue architecture during tumorigenesis remain elusive. In skin, premalignant basal cell carcinomas form 'buds', while invasive squamous cell carcinomas initiate as 'folds'. Here, using computational modelling, genetic manipulations and biophysical measurements, we identify the biophysical underpinnings and biological consequences of these tumour architectures. Cell proliferation and actomyosin contractility dominate tissue architectures in monolayer, but not multilayer, epithelia. In stratified epidermis, meanwhile, softening and enhanced remodelling of the basement membrane promote tumour budding, while stiffening of the basement membrane promotes folding. Additional key forces stem from the stratification and differentiation of progenitor cells. Tumour-specific suprabasal stiffness gradients are generated as oncogenic lesions progress towards malignancy, which we computationally predict will alter extensile tensions on the tumour basement membrane. The pathophysiologic ramifications of this prediction are profound. Genetically decreasing the stiffness of basement membranes increases membrane tensions in silico and potentiates the progression of invasive squamous cell carcinomas in vivo. Our findings suggest that mechanical forces—exerted from above and below progenitors of multilayered epithelia—function to shape premalignant tumour architectures and influence tumour progression.

Physical forces often act within defined boundaries to generate tissue shapes[2]. Tumours are a primary example of tissue growth within spatial constraints, which include neighbouring cells and extracellular matrix (ECM)[3]. Mechanical properties and forces acting on solid tumours are likely to be particularly complex, as these tumours are heterogeneous in cellular composition, and they inhabit distinct ECMs[4].

Solid tumours that initiate from stratified tissues present an opportunity to investigate the diverse physical constraints involved in tumorigenesis. In the epidermis, proliferative progenitors continually commit to terminal differentiation, exiting the inner (basal) layer and moving upward to replenish the skin's barrier[5]. Here, we focus on two common skin cancers that originate from basal epidermal progenitors. Basal cell carcinomas (BCCs), driven by constitutive activators of Sonic hedgehog signalling (for example, SmoM2), bud inward into surrounding stroma but appear to retain their basement membrane and rarely spread to neighbouring tissues[6,7]. By contrast, squamous cell carcinomas (SCCs), driven by oncogenic activators of RAS/MAPK signalling (for example, HRas^{G12V}; ref. [8]), initiate as bidirectional tissue folds before becoming invasive and aggressive. Our study unearths previously unappreciated forces from overlying suprabasal tumour cells and underlying ECM that profoundly affect tumour architecture and malignancy.

Tumour architectures of BCCs and SCCs

To explore early steps in BCC and SCC tumorigenesis, we used low-titre in utero lentiviral (LV) delivery[9] to selectively transduce Cre recombinase (LV–Cre–H2B–RFP, where H2B is histone 2B and RFP is red fluorescent protein) into the single-layered skin epithelium of embryos at day 9.5 of development (E9.5) from either *R26–SmoM2–YFP*$^{fl/fl}$ ('SmoM2') or *HRas-G12V*$^{fl/fl}$;*R26–YFP*$^{fl/+}$ ('HRas^{G12V}') mice (Fig. 1a) (where YFP is yellow fluorescent protein). By E18.5, when normal epidermal maturation is complete, early hyperplastic lesions were evident that progressed to BCCs (SmoM2) and benign papillomas or SCCs (HRas^{G12V}) in adulthood (Fig. 1b, c and Extended Data Fig. 1a)[10]. Even during these initial oncogenic stages, lesions expressing mutant SmoM2 or HRas^{G12V} displayed distinct tissue architectures.

Both SmoM2 and HRas^{G12V} lesions were displaced vertically from the epidermal plane (measured as basal indentation depth, I_B), but they had different curvature radii of the basal leading edge (denoted $\varnothing_c$; Fig. 1a, d). We describe these distinct tissue architectures by a shape factor, S, defined as the ratio of I_B to $\varnothing_c$. High S values indicate deeply invaginating and small curvature radius growths (that is, BCC-like 'buds'), while low S values indicate high curvature radii and shallow invaginations and/or evaginations (that is, SCC-like 'folds') (Fig. 1d).

---

[1]Howard Hughes Medical Institute, Robin Chemers Neustein Laboratory of Mammalian Cell Biology and Development, The Rockefeller University, New York, NY, USA. [2]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. [3]Jozef Stefan Institute, Ljubljana, Slovenia. [4]Electron Microscopy Resource Center, The Rockefeller University, New York, NY, USA. [5]Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ, USA. [6]Department of Molecular Biology, Princeton University, Princeton, NJ, USA. [7]Present address: Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, USA. ✉e-mail: fuchslb@rockefeller.edu

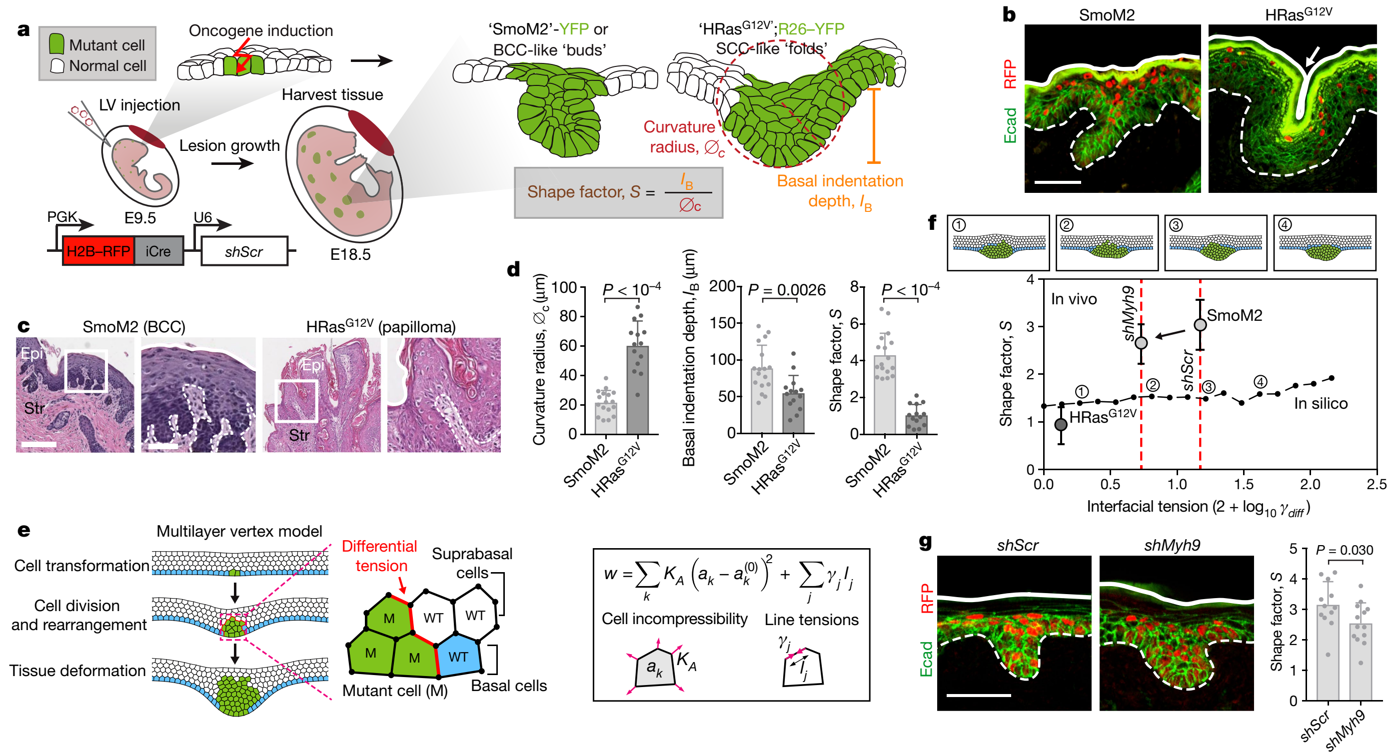


Fig. 1 | Multilayer in vivo and in silico models of early tumour morphogenesis. a, Mouse models of oncogenesis. SmoM2 or HRas^{G12V} embryos at E9.5 were transduced in utero with LV–Cre; at E18.5, tissues were harvested and tumour architectures were analysed for the indicated parameters. The lentiviral vector is shown at the bottom left. iCre, improved Cre recombinase; PGK, phosphoglycerate kinase. **b**, Immunofluorescence images of oncogenic growths. Dotted lines, epithelial–stromal borders; solid lines, apical borders; arrow, apical fold; Ecad, E-cadherin. Scale bar, 50 μm. **c**, Histology of adult mouse tumours from these mice. Epi, epithelium; Str, stroma. Scale bars: left, 250 μm; right (zoom-in), 100 μm. **d**, Quantifications of lesions (SmoM2, $n = 17$; HRas^{G12V}, $n = 14$) from four embryos (taken from two litters) per condition (means + s.d., two-tailed unpaired t-test). This experiment was independently repeated twice. **e**, Multilayer epithelium vertex model. A single basal cell is transformed (green) and then undergoes cycles of division. Differential line tensions at interfaces between mutant (M) and wild-type (WT) cells simulate interfacial tensions. Forces are contributed by cell area incompressibility (K_A) and apical, basal and lateral line tensions (γ_j); effective energy (w) is minimized (k refers to an individual cell; j is the cell edge; and a_k is the cross-sectional area). See Supplementary Note 1 for details. **f**, Bottom, effects of varying interfacial tensions (γ_{diff}) on tumour architecture. S values (median, from $n = 5$ independent simulations) from in silico modelling are plotted as a black line. Example snapshots for the indicated values of γ_I are shown at the top. Experimental data from **d**, **g** (for *shMyh9*, SmoM2 and HRas^{G12V}; mean ± s.d.) are overlaid. **g**, Left, immunofluorescence images of SmoM2 embryos transduced with LV–Cre containing either scrambled hairpin control (*shScr*) or *Myh9*-targeted shRNA (*shMyh9*). Scale bar, 50 μm. Right, lesion shape factors (*shScr*, $n = 12$; SmoM2, *shMyh9*, $n = 13$; mean + s.d., two-tailed unpaired t-test) from four embryos, two litters per condition.

HRas^{G12V} folds were further distinguished by having an invaginated apical surface (apical indentation depth, I_A). Although SmoM2 and HRas^{G12V} lesions could be distinguished by additional morphological parameters, S differentiates these phenotypes over a large range of shape variations in two and three dimensions (Extended Data Fig. 1b–d), demonstrating its utility in quantifying oncogenic tissue architectures.

Role of proliferation in architecture

As expected, proliferation was increased in all oncogenic clones, and this was evident at E15.5, before vertical tissue displacements (Extended Data Fig. 2a). Indicative of cellular crowding, oncogenic basal cells also displayed a higher cell density and more columnar shape (denoted the basolateral aspect ratio, A_B) than neighbouring wild-type cells (Extended Data Fig. 2b).

To investigate whether the increased proliferation of oncogenic basal cells within a confined epithelial space drives tissue deformations, we used LV transduction of the cell-cycle inhibitor p27^{Kip1} (*LV–H2B–RFP–TRE–Cdkn1b*, where *Cdkn1b* encodes p27^{Kip1}) to controllably decrease proliferation in developing oncogenic skin (Extended Data Fig. 2c, d). In embryos containing a basal-cell-targeted, tetracycline-inducible transactivator (*Krt14–rtTA*), p27^{Kip1} activation markedly reduced proliferation in transduced patches. This led to a dose-dependent decrease in lesion size and basal indentation in both SmoM2 and HRas^{G12V} oncogenic skins (Extended Data Fig. 2e). Thus, although proliferation provides the driving force for growth expansion and out-of-plane tissue deformation, it does not explain these distinct tumour architectures.

Role of interfacial actomyosin tension

Because actomyosin is a major biophysical driver of architecture in simple epithelia and their associated tumours[11,12], we next turned to whether differences in polarized actomyosin-driven tensions might drive differences in tumour architecture in stratified epithelia. We carried out laser ablation of cell junctions at interfaces between mutant basal cells and their neighbours in live embryos. Recoil velocities were substantially higher at interfaces between wild-type and SmoM2 cells than between wild-type and HRas^{G12V} cells (Extended Data Fig. 3a). Staining for the actomyosin contractile machinery corroborated these findings (Extended Data Fig. 3b). These data are consistent with the anisotropic and circumferentially oriented elongation along SmoM2 and wild-type cell borders, and demonstrate differential cell–cell interfacial tension (Extended Data Fig. 3c). Consistent with the differential recoil velocities, changes in actomyosin localization were not seen in HRas^{G12V} lesions.

To systematically explore the physical mechanisms underlying oncogenic tissue morphogenesis in skin, we developed a minimal mechanical model of a multilayered epithelium. We described the tissue in cross-section as a five-cell-layer-wide band with a vertex model[13] to mimic stratified epithelial architectures. To model oncogenic transformation, we induced cell proliferation to match experimentally observed cell counts, which resulted in cell deformations, rearrangements and tissue-scale shape changes (Fig. 1e, Supplementary Video 1 and Supplementary Note 1).

To probe the role of measured cell–cell interfacial tensions, we adjusted surface tensions at mutant–wild-type interfaces ($\gamma_{\text{M-WT}}$). Surprisingly, differential cell–cell tension had a minimal effect on lesion architectures in this stratified model (Fig. 1f). Predicted shapes were exclusively bud-like, and the main effect of increasing $\gamma_{\text{M-WT}}$ was to increase the compactness and reduce $\varnothing_c$, thus slightly increasing S. By contrast, varying tension in a monolayer model generated both apically and basally oriented tissue folds (Extended Data Fig. 3d).

Moreover, when we knocked down the dominant myosin II gene *Myh9* in SmoM2 mutants, or treated oncogenic skin explant cultures with an inhibitor of the actomyosin regulator ROCK, although actomyosin was markedly altered, only slight deviations in S were observed, and budding was still the dominant phenotype (Fig. 1g and Extended Data Fig. 3e). In line with our multilayered vertex modelling, these data suggest that the biophysical underpinnings of tumour architecture in stratified epidermis are distinct from those in previously studied simple epithelia[12].

Biophysical properties of basement membrane

Seeking alternative mechanisms that might affect tumour architectures in stratified tissues, we carried out transcriptional profiling of E15.5 epidermal progenitors. ‘Extracellular matrix’ and ‘collagen IV trimer’ were among the top gene ontology (GO)-term categories that were differentially upregulated (by a factor of two or more; $P < 0.05$) in SmoM2 versus HRas$^{\text{G12V}}$ progenitors (Fig. 2a and Extended Data Fig. 4a–c). Intriguingly, many of these genes (for example, *Lamb1*, *Col4a1/2*, *Nid1* and *Sparc*) encode components of the basement membrane—the specialized ECM that is directly underneath basal epidermal progenitors.

Owing to the importance of the ECM in shaping tissues during morphogenesis[14], we decided to explore how biophysical properties of the basement membrane might affect tumour shapes. We described the basement membrane as a thin elastic film coinciding with the basal side of progenitors[15]. In thin elastic films, both stretching and bending moduli—K_s and B, respectively—are proportional to the effective Young’s modulus (of the basement membrane in this case, E_{BM}), with stretching modulus K_s being dominant over bending modulus B ($K_s \gg B$). We also incorporated a timescale for basement-membrane assembly and remodelling (τ_a), which describes the rate of local adaptation of basement-membrane length to changes in cell dimensions resulting from growth and proliferation (Fig. 2b and Supplementary Notes 1, 2).

Gratifyingly, computationally simulated tissues were similar in shape to those we observed in vivo. Lowering the stiffness of the basement membrane or increasing its assembly rate ($1/\tau_a$) enhanced dermally oriented invaginations that are reminiscent of SmoM2 mutant buds, while sufficiently high stiffness values (roughly five times greater than basal cell stiffness) and/or moderate assembly rates resulted in basal and apical indentations, reminiscent of HRas$^{\text{G12V}}$ folds (Fig. 2c and Supplementary Video 2).

Importance of basement-membrane stiffness

To investigate the predictions of our model, we first characterized the mechanical properties of basement membranes ex vivo by atomic force microscopy (AFM; Fig. 2d and Extended Data Fig. 5a, b). In contrast to dermis, which displayed nonlinear and plastic deformations, basement membrane was much stiffer, with only slight nonlinear elasticity (Extended Data Fig. 5c). Providing experimental validation of our approximation of basement membrane as a Hookean elastic material over relevant timescales, these data point to the basement membrane as the dominant physical barrier underneath the epidermis.

SmoM2 basement membrane was softer than HRas$^{\text{G12V}}$ basement membrane at the distal leading edge of E18.5 buds (Fig. 2e), in agreement with our simulation prediction that softening of the basement membrane accentuates budding features. Moreover, the upper (proximal) SmoM2 basement membrane was stiffer than HRas$^{\text{G12V}}$ basement membrane, consistent with increased expression of genes encoding structural membrane components. Further reflecting an increased stiffness of the basement membrane, hemidesmosomal density was elevated in HRas$^{\text{G12V}}$ and proximal regions compared with distal tips of SmoM2 lesions. The stiffness of the basement membrane also increased from E15.5 to E18.5, indicative of membrane maturation—a change also accentuated in SmoM2 buds (Extended Data Fig. 5d, e).

To test the functional importance of basement-membrane stiffness in controlling tumour architectures, we began by transducing basal progenitors with short hairpin RNAs (shRNAs) targeting *Col4a1*, which encodes a key subunit of type IV collagen (colIV)—the predominant structural network that is responsible for the tensile load bearing properties of the basement membrane[16,17]. AFM measurements revealed that, relative to control scramble hairpins (*shScr*), *shCol4a1* skins displayed a marked decrease (more than 50%) in basement-membrane stiffness (Extended Data Fig. 5f). *Col4a1* knockdown in both oncogenic backgrounds accentuated downgrowth while reducing curvature radii, resulting in increased S values (Fig. 2f and Extended Data Fig. 5g).

Decreasing the levels of the colIV-crosslinking enzyme peroxidasin[18] (*shPxdn*) also reduced basement-membrane-stiffness, while decreasing the membrane-associated proteoglycan perlecan (*shHspg2*) increased stiffness. Notably, irrespective of oncogenic background, S increased when basement-membrane stiffness was reduced, and decreased when membrane stiffness was increased, consistent with our simulations (Fig. 2f and Extended Data Fig. 5f, g). Thus, SmoM2-driven buds were accentuated by reducing membrane stiffness, while HRas$^{\text{G12V}}$-driven folds were favoured by increasing stiffness.

Importance of basement-membrane assembly

Our model also predicted that differences in the dynamics of basement-membrane assembly would markedly affect tissue architecture. Interestingly, de novo assembly-promoting membrane components such as β1-subunit-containing laminin (LN-β1)[19] and nidogen were selectively enriched at the distal tips of SmoM2 buds (Fig. 2g and Extended Data Fig. 6a). Moreover, when fluorescently labelled laminin was added to oncogenic skin explant cultures, laminin incorporation into the native basement membrane was more than sixfold higher in SmoM2 than in HRas$^{\text{G12V}}$ mutants (Fig. 2h and Extended Data Fig. 6b).

In vivo, *Lamb1* shRNA knockdown markedly reduced basement-membrane assembly rates without altering stiffness (Extended Data Fig. 6c). In both oncogenic backgrounds, *shLamb1* decreased S, accentuating a folding architecture (Fig. 2i). Conversely, recombinant human laminin-α5β1γ1—the major de novo assembling laminin in skin and BCCs[19–21]—caused oncogenic skin explants to increase their S values and promote budding architectures (Extended Data Fig. 6d). Simulations accurately predicted these results, providing compelling evidence that rates of assembly and stiffness of basement membranes drive architectural variations (Fig. 2j).

By including biophysical properties of the basement membrane, we also more accurately simulated earlier experimental results. In particular, although cell proliferation had been predicted to have little effect on

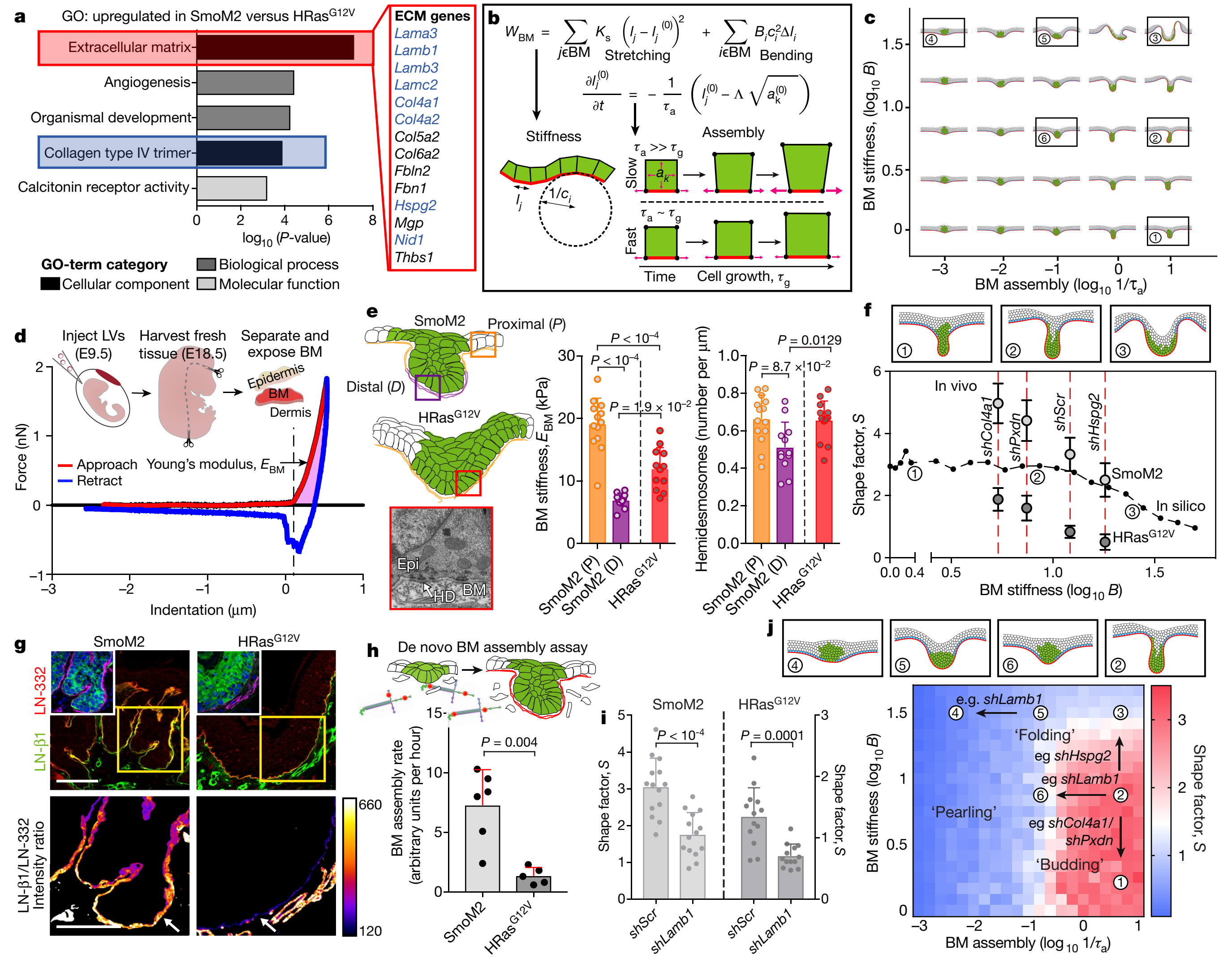


Fig. 2 | The effect of basement-membrane stiffness and assembly on tumour architectures. a, Top GO terms from E15.5 basal progenitor cell genes that are upregulated by a factor of two or more in SmoM2 versus HRas^{G12V} embryos; $n = 3$ independent biological replicates. Statistical significance was determined by unpaired, two-tailed *t*-test and *P* values were corrected using the Benjamini–Hochberg method. ECM genes encoding known basement-membrane components are in blue. **b**, Modelling the biophysical properties of the basement membrane (BM). The membrane stretching modulus (K_S), bending modulus (B) and assembly rate ($1/\tau_a$) are incorporated into the effective energy term (w_{BM}). Assembly of basement membranes is proportional to the rate constant τ_a, and cell growth is proportional to the rate constant τ_g (l_j and l_i are the lengths corresponding to cell edge j and vertex i; B is the bending modulus; C refers to curvature; see Supplementary Notes 1, 2 for details). **c**, Tissue shapes simulated by varying BM stiffnesses (proportional to B) and BM assembly rates. **d**, AFM measurements made on the BM-exposed dermal surface of EDTA-separated skin. Force-indentation curves are generated, from which the Young's modulus or stiffness of BM (E_{BM}) is calculated (see Methods). **e**, Left, diagram showing BM locations for AFM and transmission electron microscopy (TEM). Bottom left, TEM image showing electron-dense hemidesmosomes (HD) at the epidermal–dermal interface. Right, E_{BM} and ultrastructural measurements of oncogenic lesions. AFM: SmoM2 (*P*), $n = 13$; SmoM2 (*D*), $n = 11$; HRas^{G12V}, $n = 12$. TEM: SmoM2 (*P*), $n = 14$; SmoM2 (*D*), $n = 12$; HRas^{G12V}, $n = 14$. One-way analysis of variance (ANOVA) with Tukey's multiple comparisons test. **f**, Effects of varying BM stiffness on tumour architecture. *S* values (median, $n = 5$ independent simulations) from in silico modelling are plotted as a black line and overlaid with genetic data from Extended Data Fig. 5g (mean ± s.d.). E_{BM} values are indicated by red dotted lines. **g**, Immunofluorescence of laminin LN-β1, a component of nascent BMs, at the leading edge of SmoM2 and HRas^{G12V} lesions, compared with LN-332, a component of mature BMs, as shown by the intensity heatmap. Arrows mark epidermal BM. Scale bars, 50 µm. **g**, BM assembly rates measured by incorporation of fluorescent laminin into native BMs over time (SmoM2, $n = 6$ explants; HRas^{G12V}, $n = 5$ explants; two-tailed Mann–Whitney *U*-test). **i**, Quantifications of lesion *S* values following *Lamb1* knockdown. SmoM2: *shScr*, $n = 14$; *shLamb1*, $n = 14$. HRas^{G12V}: *shScr*, $n = 13$; *shLamb1*, $n = 13$. Two-tailed unpaired *t*-test; four embryos, two litters each. **j**, Comparison of experimental data and vertex model simulations for conditions in **f**, **i**. *S* values are plotted, and example simulation snapshots for the indicated values of B and $1/\tau_a$ are shown. Qualitatively distinct shape regimes include 'budding', 'folding' and 'pearling'. All bar graphs show means + s.d.

tumour shapes in the absence of basement membrane, in its presence, increased lesion deformations now surfaced (Extended Data Fig. 7a). Moreover, adding differential cell–cell tensions to the membrane mechanics model accentuated budding and expanded the diversity of tissue shapes (Extended Data Fig. 7b). Finally, although monolayer simulations accurately predicted that decreasing basement-membrane stiffness would increase *S*, they erroneously predicted that decreasing basement-membrane assembly would increase mutant basal cell crowding. Our multilayer simulations predicted a near-constant density of oncogenic progenitors, which matched that observed upon knockdown of *Lamb1* or *Col4a1* (Extended Data Fig. 7b–d). Overall, our experiments and simulations best suit a multilayered model with the presence of basement membrane, in which basal cells can transition into suprabasal layers.

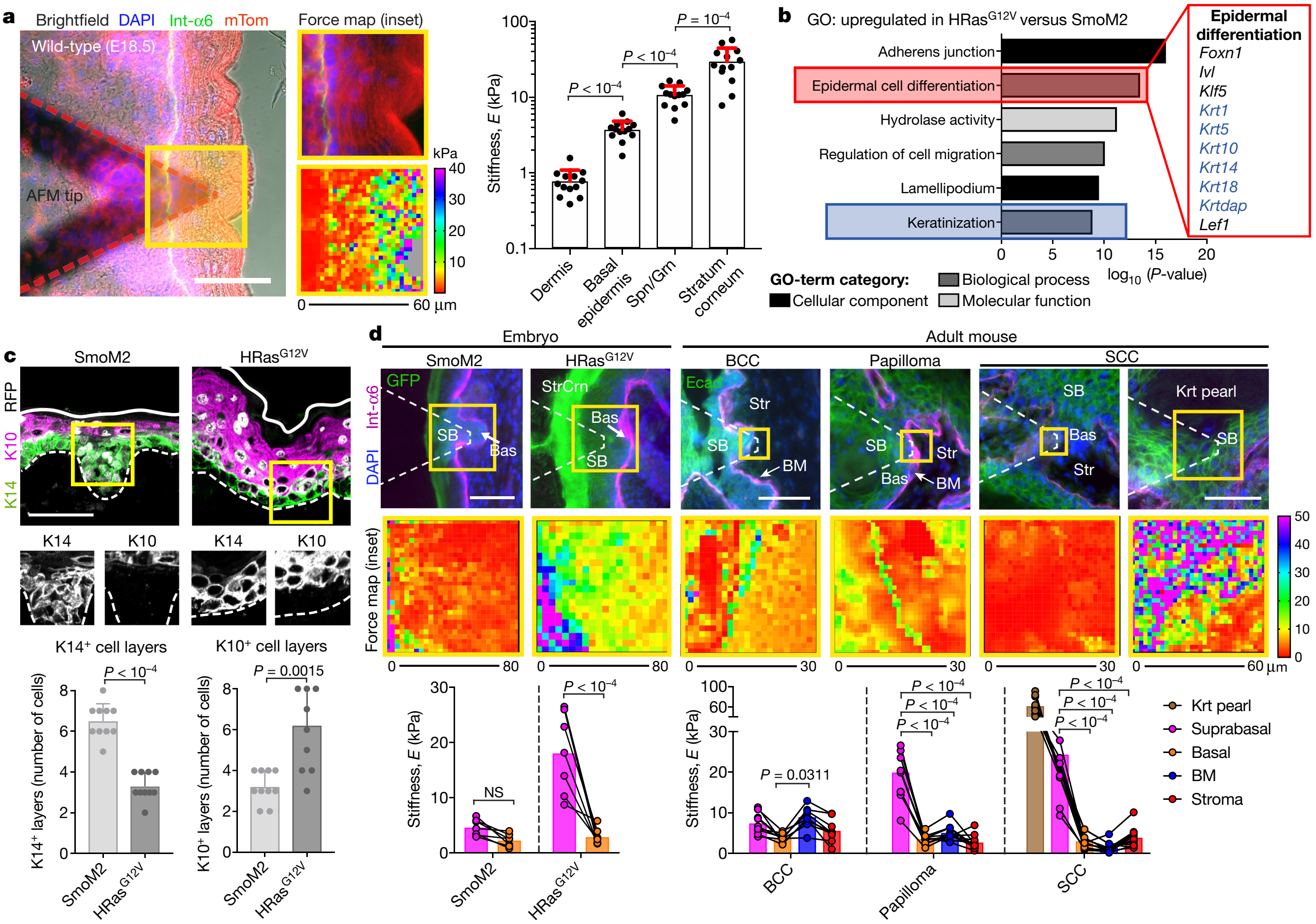


Fig. 3 | Tumour-specific subrabasal stiffness gradients. a, Left, fluorescence image of fresh E18.5 skin overlaid with an AFM tip, probed for spatially resolved stiffness measurements. DAPI stains nucleic acid; Int-a6 encodes integrin α6 and marks epithelial–stromal borders; mTom (mTomato) stains plasma membranes. Centre, force maps of the yellow-boxed region. Right, stiffness values for layers of stratified skin (dermis, basal epidermis, spinous and granular suprabasal layers (Spn/Grn), and stratum corneum; $n = 13$ regions from four embryos; mean + s.d.; Holm–Sidak's multiple comparisons test). **b**, Top GO terms for mRNAs upregulated by twofold or more in HRas^{G12V} versus SmoM2 progenitors, from $n = 3$ independent biological replicates. Keratins are shown in blue. Statistical significance was determined by unpaired two-tailed *t*-test and *P*-values were corrected using the Benjamini–Hochberg method. **c**, Top, immunofluorescence images of keratin 14 (K14)$^+$ and keratin 10 (K10)$^+$ expression, and bottom, their quantifications (means + s.d.) for SmoM2 ($n = 11$) and HRas^{G12V} ($n = 11$) lesions from three embryos each. **d**, AFM stiffness maps of oncogenic embryonic and adult SmoM2- and HRas^{G12V}-driven lesions. Suprabasal (SB), basal (Bas), BM, keratin pearl (Krt pearl) and stroma (Str) tumour regions are indicated. Paired measurements of each tumour region for SmoM2 ($n = 7$), HRas^{G12V} ($n = 7$), BCCs ($n = 9$), papillomas ($n = 9$) and SCCs ($n = 9$) lesions/tumours are shown (means; two-way ANOVA with Bonferroni's multiple comparisons test; NS, not significant). Scale bars, 50 μm.

Tumour-specific suprabasal stiffness gradients

Thus far, our simulations had treated basal and suprabasal layers analogously, except in their proliferative status. However, our multilayered epithelial simulations led us to wonder whether suprabasal cell mechanics might be an additional biophysical player in sculpting tumour architecture. We therefore turned to addressing whether the changes in gene expression that occur as basal progenitors commit to terminal differentiation[22] might affect the skin's mechanical properties, and if so, how.

We performed AFM and coincident optical imaging to measure the cell-layer-specific moduli of skin (Fig. 3a). Interestingly, the basal layer was roughly five times stiffer than papillary dermis (3 kPa versus 0.6 kPa), while spinous and granular layers, marked by keratin K10, were four times stiffer than the basal layer. Harbouring flattened layers of dead, enucleated squames, the outermost stratum corneum showed the greatest stiffness.

SmoM2 and HRas^{G12V} progenitors undergo distinct differentiation programs[23]. Our transcriptional profiling highlighted these distinctions, with the GO terms 'epidermal differentiation', 'keratinization' and 'adherens junction' being enriched in HRas^{G12V} versus SmoM2 basal cells (Fig. 3b). Correspondingly, suprabasal SmoM2 bud cells were K14$^+$ while HRas^{G12V} suprabasal layers were expanded and K10$^+$ (Fig. 3c). The mechanical properties mirrored these differences: HRas^{G12V} mutants exhibited a large stiffness difference between suprabasal and basal compartments that was nearly absent in SmoM2 mutants (Fig. 3d).

Notably, tumour architectures and differentiation-correlated stiffness gradients extended to adult mouse and human BCCs, papillomas and SCCs. BCCs exhibited only a slight elevation in stiffness (roughly 1.5-fold) from basal to suprabasal layers, while having diminished curvature radii by comparison with SCC counterparts (Fig. 3d and Extended Data Fig. 8a, b). A pronounced stiffening of the basement-membrane region was seen at the BCC–stroma interface. This correlated with RNA-sequencing data, which showed that purified basal progenitors from adult SmoM2 tumours that had invaginated into the dermal compartment (α6^{hi} YFP$^+$ Sca1^{neg}) had substantially higher expression of ECM/basement-membrane genes than those that remained within the interfollicular epidermis (α6^{hi} YFP$^+$ Sca1$^+$; Extended Data Fig. 8c, d). By contrast, benign HRas^{G12V}-driven papillomas had a modestly stiff basement

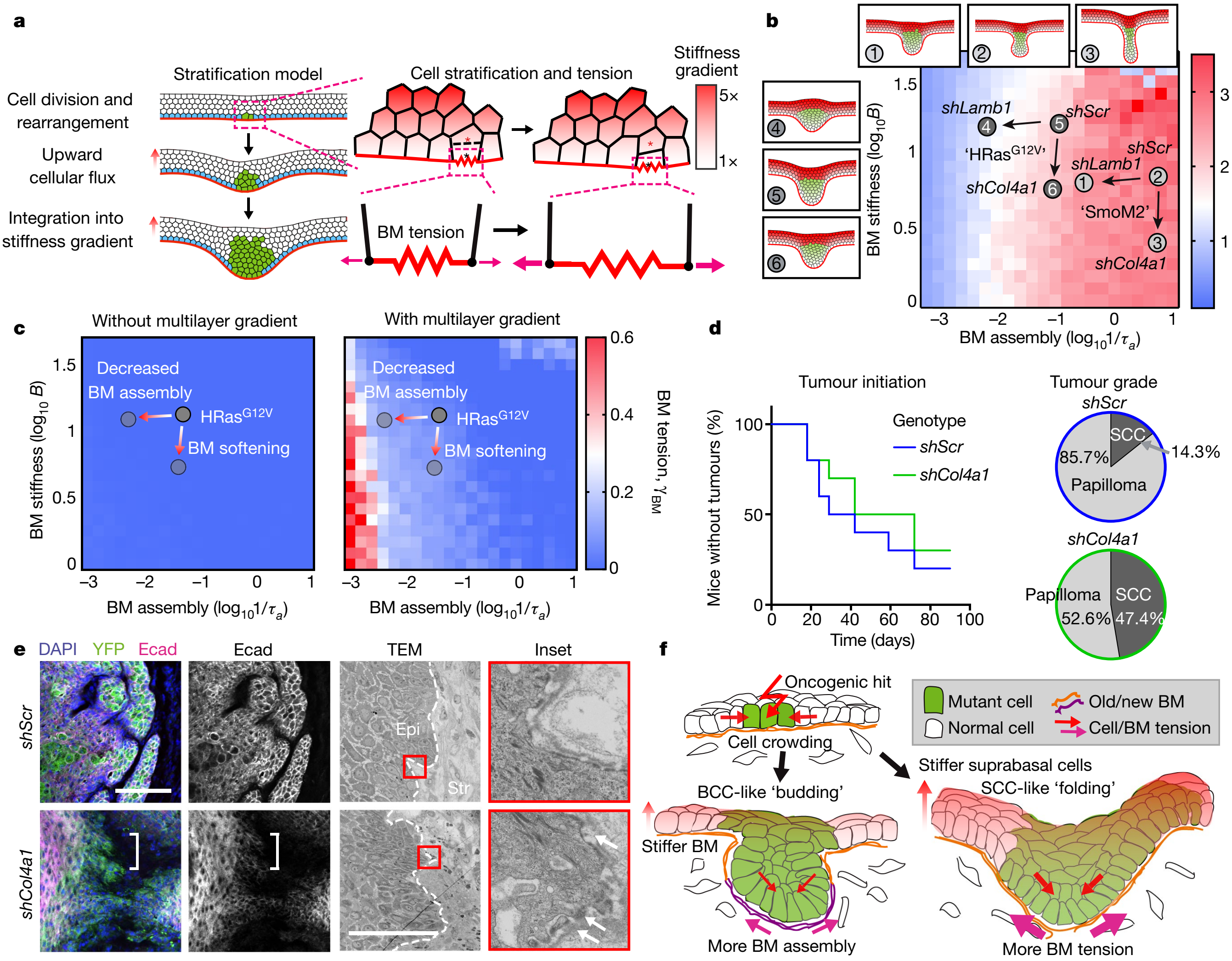


Fig. 4 | A role for the mechanics of stratified cells in tumour invasion. **a**, Multilayered epithelial vertex model. Tumour cells move upwards into suprabasal layers in a manner that depends on junctional tension and division orientation, while lateral cell tension increases as a function of vertical position. Red and black asterisks indicate a pair of dividing cells. **b**, Comparison of *S* values between experimental data and multilayer vertex model simulations that include a suprabasal stiffness gradient. *S* values are indicated by heatmap. Tissue architecture examples for experimental parameter values of *shScr*, *shLamb1* and *shCol4a1* are shown. Arrows denote changes in basement-membrane properties due to shRNAs. **c**, Changes in BM tension resulting from the multilayer stiffness gradient. Extensile tensions acting on the BM were calculated in the absence (left) and presence (right) of this gradient. **d**, Impact of decreasing BM stiffness (*shCol4a1*) on the percentage of mice bearing HRas^{G12V} tumours over time. Papillomas and SCCs were distinguished by tumour pathology upon completion of the experiment (*shScr*, $n = 10$ mice; *shCol4a1*, $n = 10$ mice). **e**, Immunofluorescence and ultrastructure imaging of *shCol4a1* versus *shScr* tumours from age-matched littermates. Regions with Ecad downregulation (left, brackets) and BM discontinuities (right, arrows) are indicated. Scale bars, 50 µm. **f**, Summary of the mechanical forces that affect tumour architecture and invasion.

membrane but substantial suprabasal stiffening. Most notable were SCCs: their hallmark keratinized pearls exhibited extraordinary stiffness (Fig. 3d). Correspondingly, and characteristic of invasive cancers, SCCs showed the lowest stiffness within the basement-membrane region compared with the other tumours.

Stratified cell mechanics and tumour invasion

Given these results, we decided to incorporate a suprabasal stiffness gradient into our multilayered simulations (Fig. 4a). We allowed progenitors to 'differentiate' and move upward into this pre-existing suprabasal stiffness gradient (Fig. 4a and Supplementary Note 3). As a consequence, tumours with high *S* shapes shifted towards higher membrane stiffness and reduced apical indentation was observed (Fig. 4b, Extended Data Fig. 9a, b and Supplementary Video 3).

To assess the functional significance of these predicted effects, we transduced embryos harbouring a suprabasal-specific involucrin promoter driving rtTA (*Inv–rtTA*) with *TRE–Klhl16*, whose encoded ubiquitin ligase causes degradation of keratin networks[24]. Doxycycline induction resulted in reduced suprabasal cell stiffness and increased I_A values in HRas^{G12V} mutants (Extended Data Fig. 9c–e).

Although the consequences of suprabasal stiffening for tumour shape were relatively modest, our model intriguingly predicted marked effects on extensile tensions of the tumour basement membrane (Fig. 4a, c). In a multilayered gradient of suprabasal stiffness, basement-membrane tension was predicted to be pronounced under conditions in which membrane-assembly rates were slow, namely in HRas^{G12V}-driven tumours (Fig. 4c). Moreover, the effects of extensile tensions were predicted to be most pronounced when the stiffness of the basement membrane was reduced and suprabasal stiffness was elevated.

To test these predictions in vivo, we knocked down *Col4a1* in HRas^{G12V} skin progenitors and monitored the effects of reducing the stiffness of the basement membrane as tumours progressed from papillomas to SCCs in adult mice. Although the incidence of papilloma formation (that is, tumour initiation) was comparable to the effects of *shScr*, *shCol4a1* greatly accelerated papilloma progression into invasive SCCs (Fig. 4d). Moreover, at the ultrastructural level, the basement membrane became considerably more discontinuous in *shCol4a1* than in *shScr* SCCs, while the tumour epithelium showed hallmarks of invasion, including spindle-shaped cell morphology and diminished E-cadherin at cell–cell borders (Fig. 4e).

Discussion

By combining computational predictions with biophysical measurements and genetic manipulations, we have systematically unearthed constraining mechanical forces that coalesce at the basement membrane to govern the architecture and behaviour of cancers originating from stratified squamous epithelia (Fig. 4f). Given the distinct material properties that can be generated by oncogene-induced changes in the stiffness and assembly of basement membranes, and also in cellular differentiation programs[2,25], the combination of these influences begins to explain the remarkable diversity in architectures of complex tissues and their cancers, and sets tumours of stratified epithelia apart from their simple epithelial counterparts[12].

Our findings are interesting in light of recent reports that the mechanics of basement membranes can influence tissue morphogenesis and invasion[26–30]. We have shown that if mechanical forces transmitted by overlying differentiated cells are sufficiently strong, as they are in SCCs, tensile stresses experienced in the underlying basement membrane may contribute to loss of membrane integrity. Our findings also suggest that once integrity is lost—for instance through tumour-induced enzymatic digestion of ECM—forces emanating from overlying differentiated tumour cells may mechanically drive the invasion of tumour-initiating progenitors at the stromal border.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2695-9.

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
2. Gilmour, D., Rembold, M. & Leptin, M. From morphogen to morphogenesis and back. *Nature* **541**, 311–320 (2017).
3. Mohammadi, H. & Sahai, E. Mechanisms and impact of altered tumour mechanics. *Nat. Cell Biol*. **20**, 766–774 (2018).
4. Pickup, M. W., Mouw, J. K. & Weaver, V. M. The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep*. **15**, 1243–1253 (2014).
5. Jones, P. H., Harper, S. & Watt, F. M. Stem cell patterning and fate in human epidermis. *Cell* **80**, 83–93 (1995).
6. Atwood, S. X., Chang, A. L. & Oro, A. E. Hedgehog pathway inhibition and the race against tumor evolution. *J. Cell Biol*. **199**, 193–197 (2012).
7. Crowson, A. N. Basal cell carcinoma: biology, morphology and clinical implications. *Mod. Pathol*. **19** (Suppl 2), S127–S147 (2006).
8. Li, S., Balmain, A. & Counter, C. M. A model for RAS mutation patterns in cancers: finding the sweet spot. *Nat. Rev. Cancer* **18**, 767–777 (2018).
9. Beronja, S., Livshits, G., Williams, S. & Fuchs, E. Rapid functional dissection of genetic networks via tissue-specific transduction and RNAi in mouse embryos. *Nat. Med*. **16**, 821–827 (2010).
10. Beronja, S. et al. RNAi screens in mice identify physiological regulators of oncogenic growth. *Nature* **501**, 185–190 (2013).
11. Munjal, A. & Lecuit, T. Actomyosin networks and tissue morphogenesis. *Development* **141**, 1789–1793 (2014).
12. Messal, H. A. et al. Tissue curvature and apicobasal mechanical tension imbalance instruct cancer morphogenesis. *Nature* **566**, 126–130 (2019).
13. Farhadifar, R., Röper, J. C., Aigouy, B., Eaton, S. & Jülicher, F. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Curr. Biol*. **17**, 2095–2104 (2007).
14. Daley, W. P. & Yamada, K. M. ECM-modulated cellular dynamics as a driving force for tissue morphogenesis. *Curr. Opin. Genet. Dev*. **23**, 408–414 (2013).
15. Štorgel, N., Krajnc, M., Mrak, P., Štrus, J. & Ziherl, P. Quantitative morphology of epithelial folds. *Biophys. J*. **110**, 269–277 (2016).
16. Pöschl, E. et al. Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development. *Development* **131**, 1619–1628 (2004).
17. Brown, K. L., Cummings, C. F., Vanacore, R. M. & Hudson, B. G. Building collagen IV smart scaffolds on the outside of cells. *Protein Sci*. **26**, 2151–2161 (2017).
18. Fidler, A. L. et al. A unique covalent bond in basement membrane is a primordial innovation for tissue evolution. *Proc. Natl Acad. Sci. USA* **111**, 331–336 (2014).
19. Yurchenco, P. D. Integrating activities of laminins that drive basement membrane assembly and function. *Curr. Top. Membr*. **76**, 1–30 (2015).
20. Li, J. et al. Laminin-10 is crucial for hair morphogenesis. *EMBO J*. **22**, 2400–2410 (2003).
21. DeRouen, M. C. et al. Laminin-511 and integrin beta-1 in hair follicle development and basal cell carcinoma formation. *BMC Dev. Biol*. **10**, 112 (2010).
22. Asare, A., Levorse, J. & Fuchs, E. Coupling organelle inheritance with mitosis to balance growth and differentiation. *Science* **355**, eaah4701 (2017).
23. Gonzales, K. A. U. & Fuchs, E. Skin and its regenerative powers: an alliance between stem cells and their niche. *Dev. Cell* **43**, 387–401 (2017).
24. Büchau, F., Munz, C., Has, C., Lehmann, R. & Magin, T. M. KLHL16 degrades epidermal keratins. *J. Invest. Dermatol*. **138**, 1871–1873 (2018).
25. Discher, D. E., Mooney, D. J. & Zandstra, P. W. Growth factors, matrices, and forces combine and control stem cells. *Science* **324**, 1673–1677 (2009).
26. Chaudhuri, O. et al. Extracellular matrix stiffness and composition jointly regulate the induction of malignant phenotypes in mammary epithelium. *Nat. Mater*. **13**, 970–978 (2014).
27. Crest, J., Diz-Muñoz, A., Chen, D. Y., Fletcher, D. A. & Bilder, D. Organ sculpting by patterned extracellular matrix stiffness. *eLife* **6**, e24958 (2017).
28. Glentis, A. et al. Cancer-associated fibroblasts induce metalloprotease-independent cancer cell invasion of the basement membrane. *Nat. Commun*. **8**, 924 (2017).
29. Harunaga, J. S., Doyle, A. D. & Yamada, K. M. Local and global dynamics of the basement membrane during branching morphogenesis require protease activity and actomyosin contractility. *Dev. Biol*. **394**, 197–205 (2014).
30. Kelley, L. C. et al. Adaptive F-Actin polymerization and localized ATP production drive basement membrane invasion in the absence of MMPs. *Dev. Cell* **48**, 313–328 (2019).

Methods

Mouse lines and lentiviral constructs

All animal experiments were performed in the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)-accredited Comparative Bioscience Center at The Rockefeller University. Experiments were performed in accordance with National Institutes of Health (NIH) guidelines for Animal Care and Use, approved and overseen by The Rockefeller University's Institutional Animal Care and Use Committee (IACUC). The following previously generated mouse lines were used here: *Rosa26–SmoM2–YFP*$^{\mathrm{fl/fl}}$ (ref.[31]), *FrHRas–G12V*$^{\mathrm{fl/fl}}$ (ref.[32]), *Rosa26–EYFP*$^{\mathrm{fl/fl}}$ (ref.[33]), *Rosa26*^{mTmG} (ref.[34]), *Krt14–rtTA* (Fuchs laboratory) and *hIVL–rtTA* (ref. [35]). C57Bl6J/CD1 mixed-background strains were used. Embryos were injected with lentivirus at 9.5 days post-coitum (dpc) as described[9]. To induce recombination of transgenic cassettes, the following lentiviruses were injected: *LV–Cre*, *LV–nls–iCreH2BRFP*, or *LV–nls–iCreH2BGFP*[9]. shRNA clones were obtained from The RNAi Consortium (TRC) shRNA library (Sigma), present in the pLKO.1-puro vector and tested for knockdown efficiency in primary mouse keratinocytes isolated as previously described[41]. These cells were not routinely tested for mycoplasma. The puro cassette was swapped out for an H2B–RFP marker before transfection into 293-FT cells for high-titre lentivirus production.

To genetically manipulate basal cell proliferation, we cloned mouse *Cdkn1b* (GenBank accession number NM009875) complementary DNA (Origene, catalogue number MR201957) into doxycycline-inducible TRE-driven pLKO.1 vectors[35] downstream of the TRE promoter using *Nhe*I/*Eco*RI restriction sites. Lentivirus was injected individually or co-injected with *LV–Cre* into *SmoM2;Krt14–rtTA*$^{+}$ mice. To genetically manipulate the stability of suprabasal cell keratin, we introduced a gene encoding a fusion of monomeric (m)RFP1 to Kelch-like protein 16 (KLHL16; Uniprot accession number Q9H2C0)[24]. Both mRFP1 and KLHL16 were assembled from Integrated DNA Technologies (IDT) gblocks and cloned into our modified pLKO.1 vector downstream of the TRE promoter using *Nhe*I/*Eco*RI restriction sites. Lentivirus was injected into *hIVL–rtTA* mice[35] or those crossed to *FrHRas–G12V*$^{\mathrm{fl/fl}}$.

shRNA sequences

Short hairpin RNA sequences were as follows—*Myh9* shRNA 1 (The RNAi Consortium (TRC) clone number (TRCN) 0000071504): 5′-CGGTAAATTCATTCGTATCAA-3′; *Myh9* shRNA 2 (TRCN0000071507): 5′-GCGATACTACTCAGGGCTTAT-3′; *Col4a1* shRNA 1 (TRCN0000311578): 5′-TCCTGGACAGGCACAAGTTAA-3′; *Col4a1* shRNA 2 (TRCN0000306536): 5′-ATCGGACCCACTGGTGATAAA-3′; *Pxdn* shRNA (TRCN0000217715): 5′-GCGGAAAGCACTAAGTGTAAA-3′; *Hspg2* shRNA (TRCN0000246981): 5′-AGCCTGACAGTGTCGAGTATA-3′; *Lamb1* shRNA 1 (TRCN 0000094314): 5′-CGCAGGTAGAAGTGAAATTAA-3′; *Lamb1* shRNA 2 (TRCN0000309482): 5-‘CGCAGGTAGAAGTGAAATTAA-3′; scramble shRNA (SHC002): 5′-CAACAAGATGAAGAGCACCAA-3′.

High-titre lentivirus production

We used 293FT cells from Thermo Fisher Scientific (catalogue number R70007). The production of vesicular stomatitis virus G (VSV-G) pseudotyped lentivirus was performed by calcium phosphate transfection of 293FT cells with pLKO plasmids and helper plasmids pMD2.G and pPAX2 (Addgene catalogue numbers 12259 and 12260). Viral supernatant was collected 46 h after transfection and filtered through a 0.45-μm filter. For in utero lentiviral transduction, viral supernatant was concentrated by ultracentrifugation. Final viral particles were resuspended in viral resuspension buffer (20 mM Tris (pH 8.0), 250 mM NaCl, 10 mM MgCl$_2$ and 5% sorbitol) and 1 μl of viral suspension was injected in utero into E9.5 embryos[9].

Immunofluorescence and antibodies

Mouse back skins were dissected and either embedded directly in optimal cutting temperature compound (OCT; premium frozen section compound, from VWR) or fixed with 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS) for 1 h at room temperature. For whole-mount imaging, embryos were fixed for 1 h in 4% paraformaldehyde, and back skin was dissected at all time points. Following fixation, samples were permeabilized in 0.3% PBS-Triton for 3–4 h at room temperature, and blocked in blocking buffer (5% donkey serum, 2.5% fish gelatin, 1% bovine serum albumin (BSA), 0.3% Triton in PBS) for 1 h at room temperature. Samples were incubated with primary antibodies at 4 °C overnight, washed for 3–4 h in PBS-Triton at room temperature, and then incubated with secondary antibodies together with 4′,6-diamidino-2-phenylindole (DAPI) overnight. Back skins were mounted in ProLong diamond antifade mountant with DAPI (Invitrogen) for imaging. For sections, back skin was placed on tissue paper, cut into strips, embedded and frozen in OCT (Leica), and sectioned with a Leica cryostat (producing sections of 12–16 μm). 5-Ethynyl-2′-deoxyuridine (EdU) was administered via intraperitoneal injection of pregnant females, which were sacrificed 30 min or 1 h post-injection; embryos were then dissected from the uterine horns. EdU labelling of embryos was performed using the Click-iT Alexa Fluor 647 Imaging kit (Thermofisher) according to the manufacturer's instructions before application of primary and secondary antibodies. Antibodies used were as follows: rat anti-RFP (Chromotek, 5F8; 1:1,000), rabbit anti-RFP (MBL, PM005; 1:1,000), chicken anti-GFP (Abcam, ab13970; 1:2,000), goat anti-P-cadherin (R&D, AF761; 1:500), rabbit anti-E-cadherin (Cell Signaling Technology, 9835; 1:500), rat anti-E-cadherin (M. Takeichi, 1:200), guinea pig anti-K14 (Fuchs laboratory; 1:500), rabbit anti-K10 (Covance, poly19054; 1:1,000), rabbit anti-collagen type IV (Abcam, ab6586; 1:500), rat anti-nidogen (Santa Cruz Biotechnology, ELM1; 1:200), rat anti-laminin-β1 (Abcam, LT3; 1:100), rabbit anti-laminin-α5 (a gift from J. Miner, Washington Univ. St Louis; 1:500), rabbit anti-laminin-332 (a gift from P. Marinkovich, Stanford Univ.; 1:500), mouse anti-phospho-S22-myosin light chain 2 (Cell Signaling Technology, 3675; 1:100), mouse anti-vimentin (Dako, 3B4; 1:200) and rat anti-Sca-1 (Becton Dickinson, D7; 1:200). All secondary antibodies used were raised in a donkey host and were conjugated to one of AlexaFluor488, AlexaFluor546 or AlexaFluor647 (Life Technologies; 1:500). Rhodamine–RRX phalloidin (Life Technologies) was used to label F-actin (1:40).

Skin explant cultures

Back skins were excised from E16.5 embryos and placed into sterile PBS. Explants were cut in half along the anterior–posterior axis to compare morphogenesis of treated versus vehicle control skin. Each explant half was placed dermis side down onto a 1.0-μm-pore-size PET Falcon cell culture insert (Becton Dickinson). Culture inserts containing skin explants were placed in prewarmed keratinocyte culture medium, and explants were kept at 37 °C, 7.5% CO$_2$ for the duration of the experiment. For actomyosin manipulation studies, 50 μM of the ROCK inhibitor Y-27632 or vehicle control (dimethylsulfoxide, DMSO) was added and samples were harvested after 24 h. For assays of basement-membrane assembly rate, laminin isolated from Engelbreth–Holm–Swarm (EHS) tumours (Millipore) was labelled with the AlexaFluor647 antibody labelling kit (A20186, ThermoFisher) according to the manufacturer's instructions, or rhodamine-labelled laminin was purchased (LMN01-A, Cytoskeleton Inc). Labelled laminin or vehicle control (PBS) was then added to explant cultures at 5 μg ml^{-1}. After 2 h, 4 h, 8 h or 16 h in culture, tissues were embedded in OCT blocks and prepared for immunofluorescence staining of the endogenous basement-membrane markers nidogen and LN-322. For gain-of-function laminin experiments, recombinant human LN-511 (BioLamina) was added at 100 μg ml^{-1} and explants were cultured for 24 h before fixation, OCT embedding, and immunofluorescence staining.

Microscopy

Confocal images were acquired using a spinning disk confocal system (Andor Technology) equipped with an Andor Zyla 4.2 camera and

Yokogawa CSU-W1 (Yokogawa Electric, Tokyo) spinning disk head on a Nikon TE2000-E inverted microscope base. Four laser lines (405 nm, 488 nm, 561 nm and 625 nm) were used for near-simultaneous excitation with a ×40/1.3 numerical aperture (NA) CFI Plan Fluor oil objective. The system was driven by Andor IQ3 software. Images of cryosections were acquired using a Zeiss Axio Observer.Z1 epifluorescent/brightfield microscope with a Hamamatsu ORCA-ER camera and an ApoTome.2 slider (to reduce light scatter in the *z* direction), controlled by ZEN Blue (Carl Zeiss, Inc.) software. All images were assembled and processed using Fiji (NIH), CellProfiler (Broad Institute) and Imaris (Oxford Instruments).

Laser ablation

Junctional laser ablations were performed on an inverted LSM 880 NLO laser scanning confocal and multiphoton microscope (Zeiss) system using a tunable Ti:sapphire near-infrared laser (Chameleon Ultra II, Coherent Scientific) tuned to 800 nm, similar to the system described in ref. [36]. Laser power and dwell time were calibrated per experiment, but power was typically between 80% and 100% transmission at a scan speed of six or five repetitions (a dwell time of 90–140 µs). Quantification of the effects of ablation was performed by manually tracing the displacement of neighbouring tricellular junctions every two frames. Instantaneous retraction velocity was measured by linear fitting of junction displacement immediately following laser ablation and calculation of the slope[37].

Image processing and analysis

Quantification of cell proliferation. Proliferation was inferred from the incorporation of labelled nucleotide analogues following a 1 h EdU pulse. EdU^+ and total basal cell nuclei were identified and counted manually on the basis of EdU and DAPI signals, respectively. Keratin 14 or P-cadherin staining was used to verify that EdU^+ cells could be found within the basal layer. The total number of EdU^+ cells was then plotted as a fraction of the total number of RFP^+ basal cells. Measurements were pooled between multiple animals of the same genotype and used to perform unpaired analyses.

Quantification of tissue and cell morphology. Multichannel immunofluorescence images were imported into CellProfiler, and maximum projection images of small (10–14 µm) *z*-stacks were assembled. The region of epidermal tissue was identified using an adaptive Otsu thresholding strategy based on E-cadherin or keratin 14 staining. The object region of interest comprising the oncogenic lesion was then identified by H2B–RFP staining and manual selection. Rolling circles were fit to the basal-most lesion surface, from which curvature radii ($\varnothing_c$) were calculated. Ferret diameters, defined by two lines tangential to the lateral lesion edges and perpendicular to the basal layer, were calculated. A straight line perpendicular from the basal layer to the dermal-most tip of the lesion measured basal indentation depth (I_B), and shape factors (*S*) were calculated according to the equation in Extended Data Fig. 1b. Oncogenic cells were classified on the basis of the H2B–RFP signal, and the length of the basement-membrane interface was identified by α6 integrin or LN-332 staining, or manually drawn. Measurement of cell area and elongation was performed on whole-mount confocal images. Cells were segmented on the basis of cortical E-cadherin staining using a watershed algorithm. Cell elongation is defined as the ratio of major and minor axes of automatically segmented cells.

Quantification of basement-membrane assembly. Multichannel images were imported into CellProfiler, and, after background subtraction, adaptive Otsu thresholding was used to identify and mask the endogenous basement membrane on the basis of the LN-α5 immunofluorescence signal. Fluorescence signals from AlexaFluor647-labelled LN (AF647-LN) were then measured within the endogenous basement-membrane mask, and a ratiometric intensity value for AF647-LN to LN-α5 signals was calculated on a per-pixel basis. Ratiometric intensity values were calculated over 2 h, 4 h, 8 h and 16 h culture times, and a linear regression was applied to the data, from which the slope was determined. This slope gave the basement-membrane assembly rate (in fluorescence units per hour).

Atomic force microscopy

Tissue preparation for AFM measurements. To prepare skin for measurements of basement-membrane stiffness, we excised backskin at E18.5 and incubated it in 50 mM EDTA (EDTA)/PBS at 37 °C for 30 min. The epidermis and dermis were manually separated and fixed with 4% PFA for 1 h at room temperature to verify separation and lentivirus infection efficiency by optical microscopy, or the dermis was prepared directly for AFM. The dermis with basement membrane side up was affixed to a glass-bottom Petri dish using a small volume (5–8 µl) of Matrigel, after which samples were maintained in PBS with cOmplete protease-inhibitor cocktail (Roche) for the duration of the experiment. For adult tumours, freshly excised tumours were flash-frozen in OCT, and 20-µm-thick cryosections were generated. Tissue was affixed to poly-D-lysine-coated coverglass and stained for E-cadherin (Cell Signaling Technology, 9835; 1:200), α6 integrin (clone GoH3, Biolegend; 1:200) or nidogen (Santa Cruz Biotechnology, ELM1; 1:200) with AlexaFluor546 or AlexaFluor647 secondary antibodies (Life Technologies; 1:500) and Hoechst (Invitrogen; 1:1,000). All staining and incubations were carried out in PBS with 5% donkey serum with cOmplete protease-inhibitor cocktail (AFM media).

AFM measurements. A Zeiss Axio Observer inverted optical microscope (Zeiss) equipped with an MFP-3D AFM (Asylum Research) was used for all AFM experiments. AFM nanoindentation tests were performed using a 5-µm-diameter spherical tipped silicon nitride cantilever (Novascan) in AFM media. Cantilever spring constants were measured before sample analysis using the thermal fluctuation method, with nominal values of 100 pN nm^{-1}. During measurements, samples were maintained in AFM media. Brightfield, nuclei (Hoechst), E-cadherin and α6 integrin staining were captured using standard DAPI/fluorescein isothiocyanate (FITC)/tetramethylrhodamine isothiocyanate (TRITC) filter cubes and used to align the cantilever to the sample and for image co-registration. Two-dimensional force maps were taken in 20 µm × 20 µm, 30 µm × 30 µm, or 60 µm × 60 µm square grids with 20–32 sample points per axial dimension. AFM measurements were made using a cantilever deflection set point of 2 nN and an indentation rate of 22 µm s^{-1} to capture elastic properties and minimize viscoelastic effects. In all experiments, the deflection of the cantilever did not exceed the linearity of the photodiode detector, even for forces up to 10 nN. The first 100 nm of indentation were used to measure elasticity from the basement membrane. Force-indentation curves were analysed using a modified Hertz model for contact mechanics of spherical elastic bodies. The sample Poisson's ratio was assumed to be 0.4, and a power law of 1.5 was used to model tip geometry, as described[38]. To obtain Young's modulus, we equate force-indentation curves according to Equation (1), where *P* is the loading force, δ is the indentation into the material, and *R* is the effective tip curvature radius:

$$\delta = \left(\frac{9P^2}{16RE^{*2}}\right)^{1/3} \tag{1}$$

E^* is the apparent Young's modulus, defined as $\frac{1}{E^*} = \frac{1-\upsilon_1{}^2}{E_1} + \frac{1-\upsilon_2{}^2}{E_2}$, where ν_1 and ν_2 are Poisson's ratio and the subscripts denote the two contacting bodies (namely the AFM tip and the sample, respectively). For all samples tested, the value of δ at which the linear–nonlinear regime transition, or δ_L, occurred was between 0.5 nN and 1 nN, and force curves for reporting Young's moduli were fit within the linear regime. To obtain the pointwise Young's modulus, we followed the methodology of refs. [39,40]. Briefly, each data point (P_i, δ_i) in the force-indentation curve (where P_i is the loading force and δ_i is the indentation into the material) was substituted into Equation (1) to calculate the corresponding E_i

(the subscript '*i*' denotes an individual data point along the force-indentation profile). We used 1 nN as the nominal value for δ_L and calculated a Young's modulus of 70–90% of the loading curve (approximately 2 nN maximum load) for E_{high} and 10–30% of the loading curve for E_{low}. We defined an elasticity metric, $L = E_{low}/E_{high}$, where $L = 1$ is absolute linear elasticity and values of less than one are increasingly nonlinear. For plasticity measurements, we measured the difference in indentation lengths at zero force values between the approach and retraction curves. We also performed creep tests, measuring the change in indentation depth over time under constant force load, which gave qualitatively similar results for basement membrane and dermis. For adult tumour samples used for AFM analysis, serial cryosections were fixed in 4% PFA and processed for histology and immunofluorescence to verify tumour stages.

Electron microscopy

For electron microscopy, samples were fixed in 2% glutaraldehyde, 4% PFA, 1% tannic acid and 2 mM $CaCl_2$ in 0.1 M sodium cacodylate buffer, pH 7.2, at room temperature for more than 1 h, post-fixed in 1% osmium tetroxide, and processed for Epon embedding; ultrathin sections (60–65 nm) were counterstained with uranyl acetate and lead citrate. Electron-microscopy images were taken with a transmission electron microscope (Tecnai G2-12; FEI) equipped with a digital camera (AMT BioSprint29).

Fluorescence-activated cell sorting

Single-cell suspensions were obtained from either E15.5 or adult skins using published methods[41,42]. Fluorescence-activated cell sorting (FACS) was carried out using a FACSAriaII (Becton Dickinson) by The Rockefeller University FACS core facility. CD45 (biotinylated rat anti-CD45, BD Biolegend; 1:200), CD117 (biotinylated rat anti-CD117/c-kit, Biolegend; 1:200), CD31 (biotinylated rat anti-CD31/PECAM, Bioscience; 1:200), and CD140a (biotinylated rat anti-CD140a, Biolegend; 1:200) were used as lineage-negative markers (to exclude immune cells, melanoblasts, endothelium and fibroblasts, respectively). All lineage-negative cells were detected with a strepdavidin-conjugated APC/Cy7 secondary antibody (Biolegend; 1:1,000). Cells highly expressing integrin α6 (rat anti-CD49f/α6-PE, clone GoH3, Biolegend; 1:1,000) were sorted to obtain basal cells. In adult SmoM2 mice, Sca-1 (rat anti-Sca-1-PE/Cy7, clone D7, eBioscience; 1:200) was used to isolate oncogenic budded cells (Sca-1^{neg}) from oncogenic cells that remained in the epidermal layer (Sca-1^+), and CD34 (rat anti-CD34-eFluor660, clone RAM34, eBioscience; 1:200) was used to remove hair-follicle stem cells. Each sample submitted for RNA-sequencing comprised cells from at least three embryos per genotype. Cells were sorted directly into Trizol.

RNA-sequencing and RT–PCR

Total RNA was purified using a Direct-zol RNA Miniprep Plus kit (Zymo Research). Briefly, after adding 500 μl of 100% ethanol to samples, the lysate was loaded to an RNA-binding column. The column was treated with DNase I for 15 min at room temperature. After several washing steps, the RNA was eluted in DNase/RNase-free water. The quality of RNA samples was determined using an Agilent 2100 Bioanalyzer, and all samples for sequencing had RNA integrity (RIN) numbers of more than 9. Poly(A) selection and library preparation using an Illumina TrueSeq mRNA sample preparation kit, and sequencing on an Illumina HiSeq 2500 or HiSeq 4000 machine, were carried out by the Weill-Cornell Medical College Genomic Core facility. Fifty-base-pair single-end and paired-end FASTQ sequences were aligned to the mouse genome (GRCm38/mm10 annotation) using STAR (v2.6.2a)[43], and transcripts were annotated using Gencode release M9. Differential gene expression analysis was performed on the STAR gene-counts output using the DESeq2 (v1.24.0)[44] package with default parameters in RStudio (v1.1.442). Genes with a fold change of more than 2 and false discovery rate (FDR) of less than 0.1 were considered to be differentially expressed. Gene ontology terms were called using DAVID[45]. For real-time quantitative reverse transcription with polymerase chain reaction (qRT–PCR), equivalent amounts of RNA were reverse-transcribed using the SuperScript VILO cDNA synthesis kit (Invitrogen). Complementary DNAs were normalized to equal amounts using primers against *Gapdh* or *Ppib2*. cDNAs were mixed with the indicated primers and Power SYBR green PCR master mix (Applied Biosystems), and qPCR was performed using an Applied Biosystems 7900HT fast real-time PCR system. cDNAs were normalized to equal amounts using primers against *Ppib*.

Adult tumour progression studies

Embryos were injected with LV–Cre containing either scrambled control (*shScr*) or *shCol4a1* RNAs at 9.5 dpc. Up to five mice were housed per cage, with a 12-h light/dark cycle, and were provided with food and water ad libitum. Mouse experiments were performed on age-matched and strain-matched littermates randomly assigned to experimental groups. For analysis of adult tumours, tumour burden was visually inspected every two days throughout the course of the experiment, and tumour size was measured using digital calipers. Tumours were not allowed to progress beyond 2 cm in diameter, and ulceration did not exceed 10 mm in diameter, as approved by the Rockefeller University IACUC (protocol 17091-H). Ulcerations or tumours approaching these sizes were considered an end point, and the experiment was terminated at the end of three months. Tumours were excised and prepared for histology and immunofluorescence, and the number of papillomas and SCCs was assessed on the basis of histopathology.

Human research participants

De-identified, OCT-embedded fresh tissue sections of SCCs, BCCs or healthy skin from individuals that underwent Mohs micrographic surgery were used. This study did not involve the recruitment of new patients. De-identified tissue blocks were obtained from the Department of Dermatology, Weill-Cornell Medical College (New York, US). We have complied with all relevant ethical regulations: informed consent was obtained from patients by Weill-Cornell; The Rockefeller University IRB approved the use of de-identified human samples (EFU-0529).

Computational modelling

Code was written in C/C++ languages, based on standard C libraries and the GNU Scientific Library (GSL). Each simulation was run at five different seeds for random number generation, and results were averaged over these five runs. To ensure reproducibility of the results, we describe all details of the model, together with the values of model parameters, in the Supplementary Information.

Statistics and study design

In general, all experiments were repeated using at least two litters per experiment. All data sets generated were tested for normal distribution using Prism 7 (Graphpad), and all data sets that failed this test were subject to nonparametric tests for further analysis. All statistical tests performed are indicated in the figure legends. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment, except where stated.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All RNA-sequencing data from this study have been deposited in the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under accession code GSE152488 (super-series). All other data in the manuscript, supplementary materials, source data and custom code are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

Custom code for the multilayer vertex model is available upon request from M.K. (matej.krajnc@ijs.si), along with discussion/guidance for its use.

31. Mao, J. et al. A novel somatic mouse model to survey tumorigenic potential applied to the Hedgehog pathway. *Cancer Res*. **66**, 10171–10178 (2006).
32. Chen, X. et al. Endogenous expression of Hras^{G12V} induces developmental defects and neoplasms with copy number imbalances of the oncogene. *Proc. Natl Acad. Sci. USA* **106**, 7979–7984 (2009).
33. Srinivas, S. et al. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol*. **1**, 4 (2001).
34. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
35. Quiroz, F. G. et al. Liquid-liquid phase separation drives skin barrier formation. *Science* **367**, eaax9554 (2020).
36. Luxenburg, C. et al. Wdr1-mediated cell shape dynamics and cortical tension are essential for epidermal planar cell polarity. *Nat. Cell Biol*. **17**, 592–604 (2015).
37. Rauzi, M. & Lenne, P. F. Probing cell mechanics with subcellular laser dissection of actomyosin networks in the early developing *Drosophila* embryo. *Methods Mol. Biol*. **1189**, 209–218 (2015).
38. Brown, A. C., Fiore, V. F., Sulchek, T. A. & Barker, T. H. Physical and chemical microenvironmental cues orthogonally control the degree and duration of fibrosis-associated epithelial-to-mesenchymal transitions. *J. Pathol*. **229**, 25–35 (2013).
39. Costa, K. D., Sim, A. J. & Yin, F. C. Non-Hertzian approach to analyzing mechanical properties of endothelial cells probed by atomic force microscopy. *J. Biomech. Eng*. **128**, 176–184 (2006).
40. Fiore, V. F. et al. αvβ3 Integrin drives fibroblast contraction and strain stiffening of soft provisional matrix during progressive fibrosis. *JCI Insight* **3**, e97597 (2018).
41. Nowak, J. A. & Fuchs, E. Isolation and culture of epithelial stem cells. *Methods Mol. Biol*. **482**, 215–232 (2009).
42. Ouspenskaia, T., Matos, I., Mertz, A. F., Fiore, V. F. & Fuchs, E. WNT-SHH antagonism specifies and expands stem cells prior to niche formation. *Cell* **164**, 156–169 (2016).
43. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
44. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**, 550 (2014).
45. Huang, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. **37**, 1–13 (2009).

Acknowledgements We thank I. Matos, A. Asare, B. Hurwitz, S. Yuan, L. Polak, L. Hidalgo and M. Sribour for discussions and/or assistance; Y. Rominey in Memorial Sloan Kettering's Molecular Cytology core for assistance with AFM; and Rockefeller University's shared resources: the Bio-Imaging Center for microscope usage, and the Comparative Bioscience Center (AAALAC-accredited) for mouse care in accordance with NIH guidelines. V.F.F. was supported by the NIH–National Cancer Institute (NCI) Cancer Biology Training Program (grant CA009673-39) and a Charles H. Revson Senior Fellowship in Biomedical Sciences (Revson Foundation). M.K. was supported by the Slovenian Research Agency (research project Z1-1851). F.G.Q. holds a Career Award at the Scientific Interface from Burroughs Wellcome Fund. E.F. is a Howard Hughes Medical Institute (HHMI) Investigator. This research was supported by NIH grant R01-AR27883 to E.F.

Author contributions V.F.F. and E.F. conceived the experiments and wrote the manuscript, with contributions from M.K. and F.G.Q. M.K. and S.Y.S. developed the mathematical modelling. F.G.Q. designed and cloned transgenic constructs. J.L. performed lentiviral injections. H.A.P. performed ultrastructural analysis. V.F.F. performed all remaining experiments, data analyses and quantifications.

Competing interests The authors declare no competing interests.

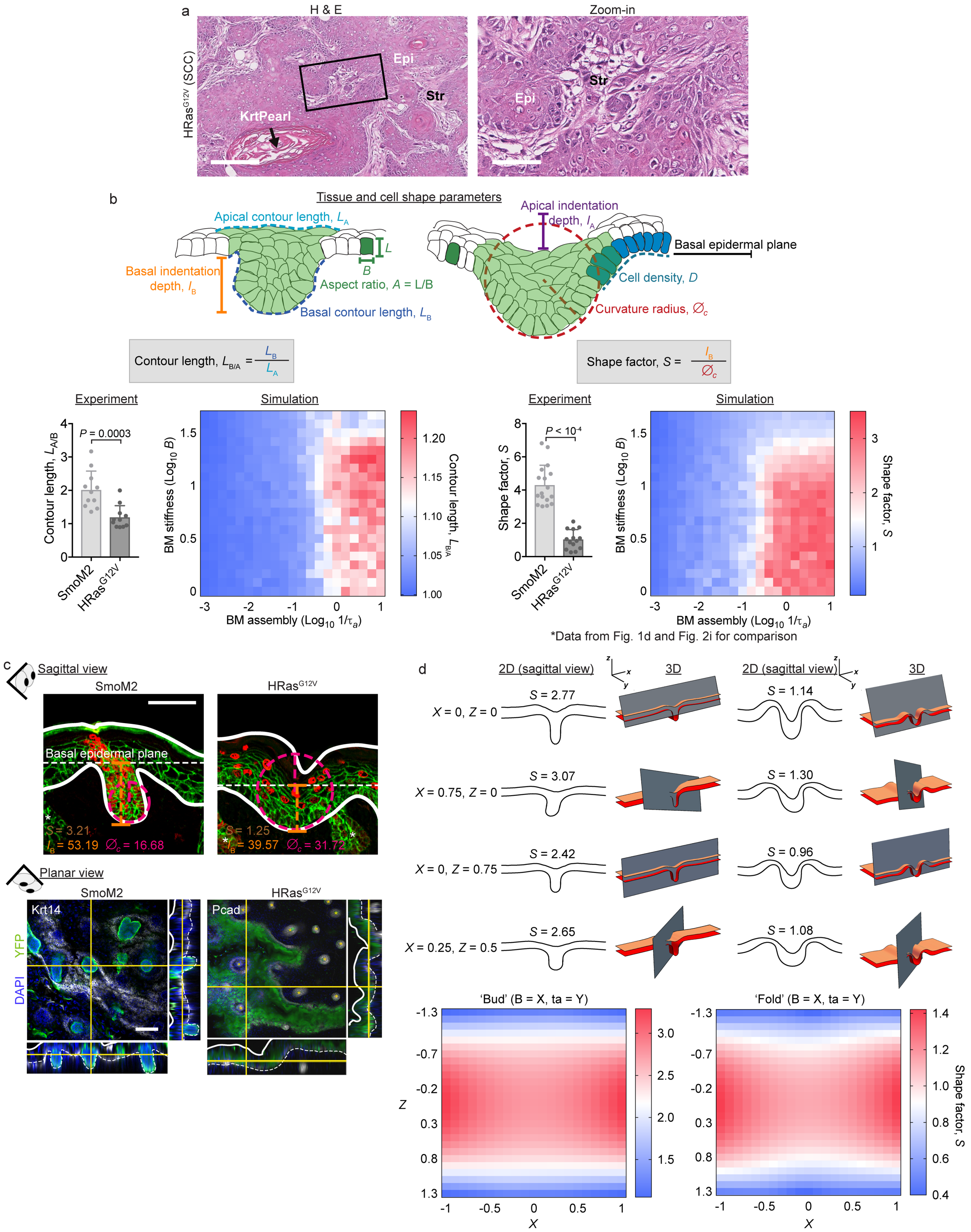
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2695-9.

Correspondence and requests for materials should be addressed to E.F.

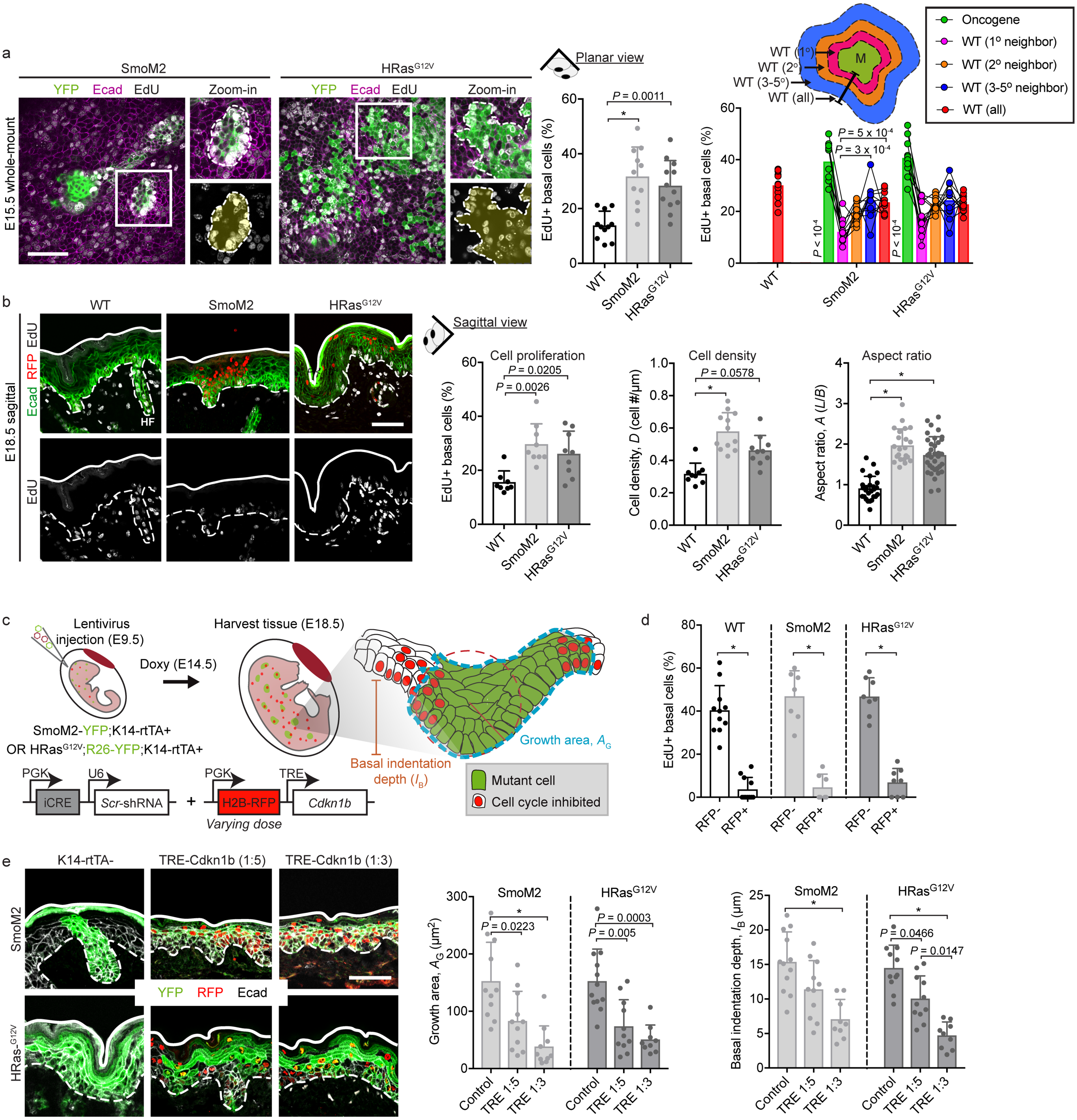
Peer review information *Nature* thanks Salvador Benitah, Nicolas Minc and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.



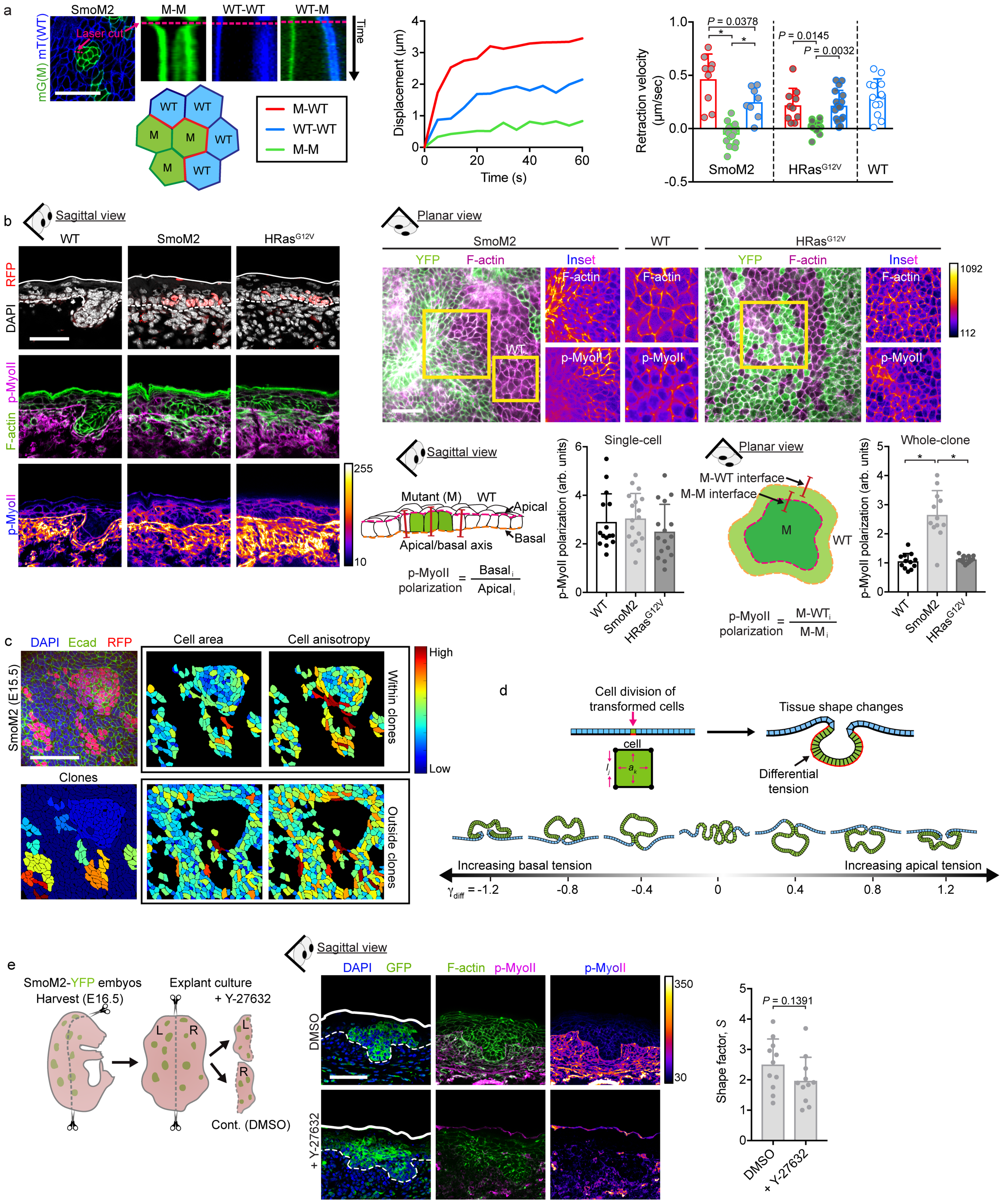
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Extended characterization of oncogenic tissue architecture models. a, Characterization of adult SCCs. E9.5 oncogenic embryos were infected in utero with LV–Cre to selectively transduce single-layered embryonic epidermis. Tissues were harvested at three months ($\mathrm{HRas}^{\mathrm{G12V}}$ SCCs) and stained with haematoxylin and eosin (H & E). Epi, epithelium; Krt pearl, keratin pearl, a hallmark of SCCs; Str, stroma. Scale bars: left, 250 µm; right (zoom-in), 100 µm. **b**, Extended description of premalignant architectures and parameters used to quantify them. The schematics at the top show all parameters used here to quantify tissue and cell-shape parameters, including apical indentation depth (I_A), basal indentation depth (I_B), apical contour length (L_A), basal contour length (L_B), curvature radius ($\varnothing_c$), cell density (D) and cell aspect ratio (A). Bottom, quantification of S values (data repeated from Fig. 1d, 2i and contour length ($L_{B/A}$; SmoM2, $n = 11$; $\mathrm{HRas}^{\mathrm{G12V}}$, $n = 11$; mean + s.d.; Mann–Whitney U-test) for lesions from four embryos, two litters for each condition. We compare experimental measurements and simulation results (see Supplementary Note 1 for modelling details), which show strong agreement. However, we note that S is better able than $L_{B/A}$ to discriminate SmoM2 and $\mathrm{HRas}^{\mathrm{G12V}}$ phenotypes. **c**, Sagittal sections and whole-mount (planar) views show the distinct tissue shapes of SmoM2 and $\mathrm{HRas}^{\mathrm{G12V}}$ lesions. Measurements of I_B and $\varnothing_c$, from which S are calculated, are depicted on example images (sagittal view). **d**, Two-dimensional (2D) and 3D simulations of tissue shapes. Archetypal budded and folded tissue architectures were simulated in 3D and cut into 2D planes with varying cutting angles X and Z (see Supplementary Note 4 for details). The resultant tissues and their calculated S values are shown. Note that both architectures are equally well discerned without systematic bias (see the range of S values). Scale bars, 50 µm.



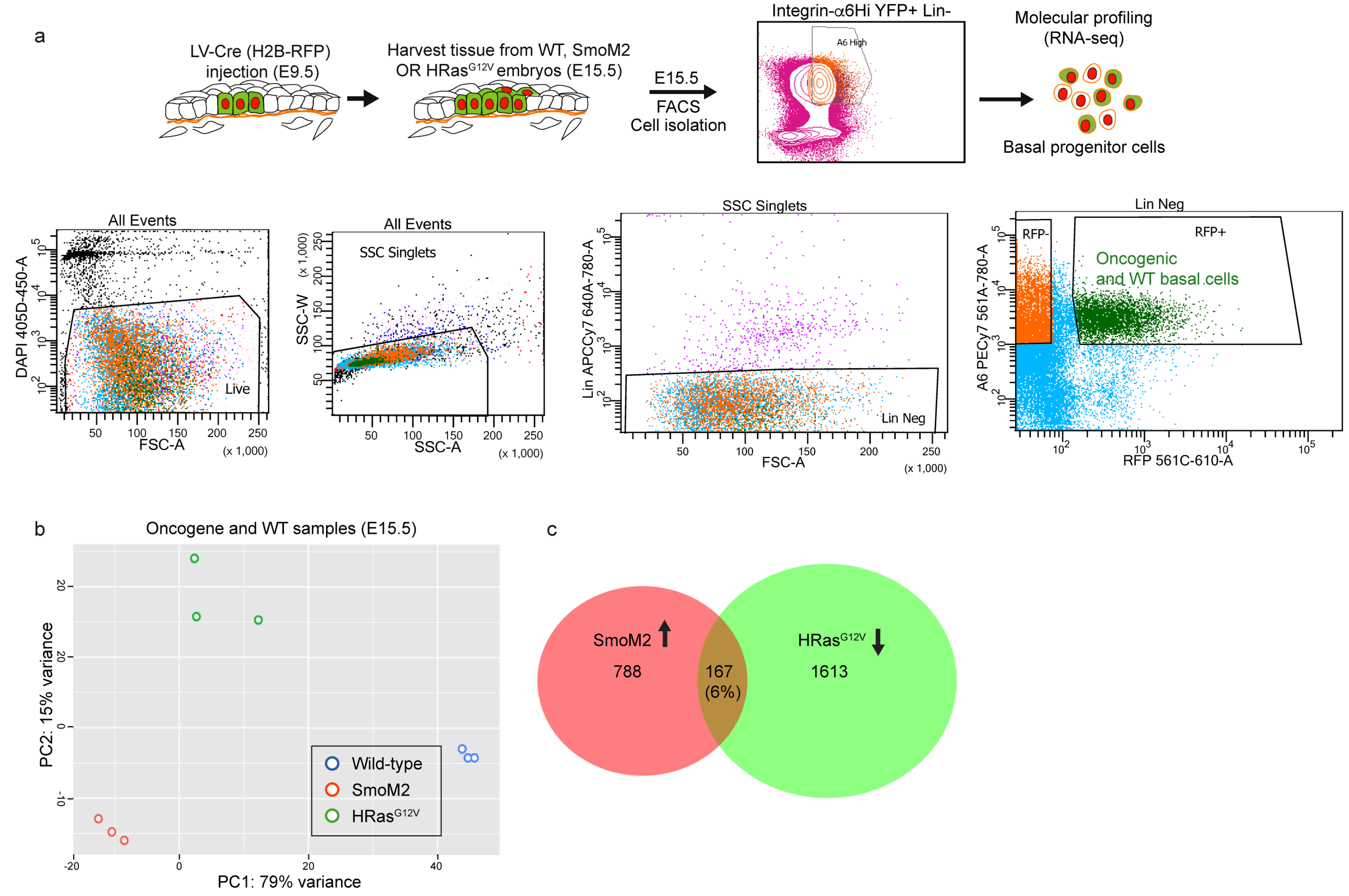
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Cell proliferation drives skin tumour growth but not architectural differences between SmoM2 and HRas^{G12V} lesions. a, Cell proliferation as measured by EdU incorporation at E15.5. SmoM2- and HRas^{G12V}-induced lesions are marked by YFP (left panels), and cell proliferation was quantified as the percentage of EdU^{+} basal cells in WT versus mutant lesions (centre panel; WT, $n = 11$; SmoM2, $n = 11$; HRas^{G12V}, $n = 12$; one-way ANOVA with Tukey's multiple comparisons test). The spatial distribution of WT cell proliferation in proximity to mutant clones was measured by quantifying the number EdU^{+} basal cells as a function of their neighbour distance from mutant clone edges (right panel; paired measurements from depicted tissue compartments; two-way ANOVA with Tukey's multiple comparisons test). HF, hair follicle. **b**, Cell proliferation as measured by EdU incorporation at E18.5. SmoM2- and HRas^{G12V}-induced lesions are marked by H2B–RFP. The graphs show cell proliferation, cell density (D) and aspect ratio (A; depicted in Extended Data Fig. 1b) at E18.5. Cell proliferation (WT, $n = 8$; SmoM2, $n = 9$; HRas^{G12V}, $n = 9$; Kruskal–Wallis test with Dunn's multiple comparisons test), D (WT, $n = 9$; SmoM2, $n = 11$; HRas^{G12V}, $n = 10$; Kruskal–Wallis test with Dunn's multiple comparisons test) and A (WT, $n = 26$; SmoM2, $n = 22$; HRas^{G12V}, $n = 36$ cells; one-way ANOVA with Tukey's multiple comparisons test) were measured for lesions from four embryos, two litters for each condition. **c**, Schematic showing our experimental approach to manipulating cell proliferation. LVs encoding H2B–GFP–iCRE and H2B–RFP–*Cdkn1b* under the control of a tetracycline-response element (TRE–*Cdkn1b*) were injected into SmoM2;K14–rtTA^{+} and HRas^{G12V};K14–rtTA^{+} mice. Embryos were injected with varying titres of TRE–*Cdkn1b* at E9.5, treated with doxycycline at E15.5, and harvested at E18.5. **d**, Cell-cycle manipulation was validated by measuring the EdU^{+} TRE–*Cdkn1b*$^{+}$ (RFP^{+}) and RFP^{-} cells in both WT and oncogenic mutant backgrounds. **e**, Immunofluorescence and quantification of oncogenic tissue architectures in oncogenic mutant embryos infected with TRE–*Cdkn1b*. Quantification of lesion growth area (A_G) and basal indentation depth (I_B) shows that lesion size and deformations decrease with an increased titre of TRE–*Cdkn1b* similarly in SmoM2 (K14–rtTA–control, $n = 11$; 1:5 TRE–*Cdkn1b*, $n = 11$; 1:3 TRE–*Cdkn1b*, $n = 10$) and HRas^{G12V} (K14–rtTA–control, $n = 12$; 1:5 TRE–*Cdkn1b*, $n = 11$; 1:3 TRE–*Cdkn1b*, $n = 9$) mutants from five embryos, two litters for each condition (one-way ANOVA with Tukey's multiple comparisons test). All bar graphs show means + s.d. Scale bars, 50 µm. $^{*}P < 0.0001$.

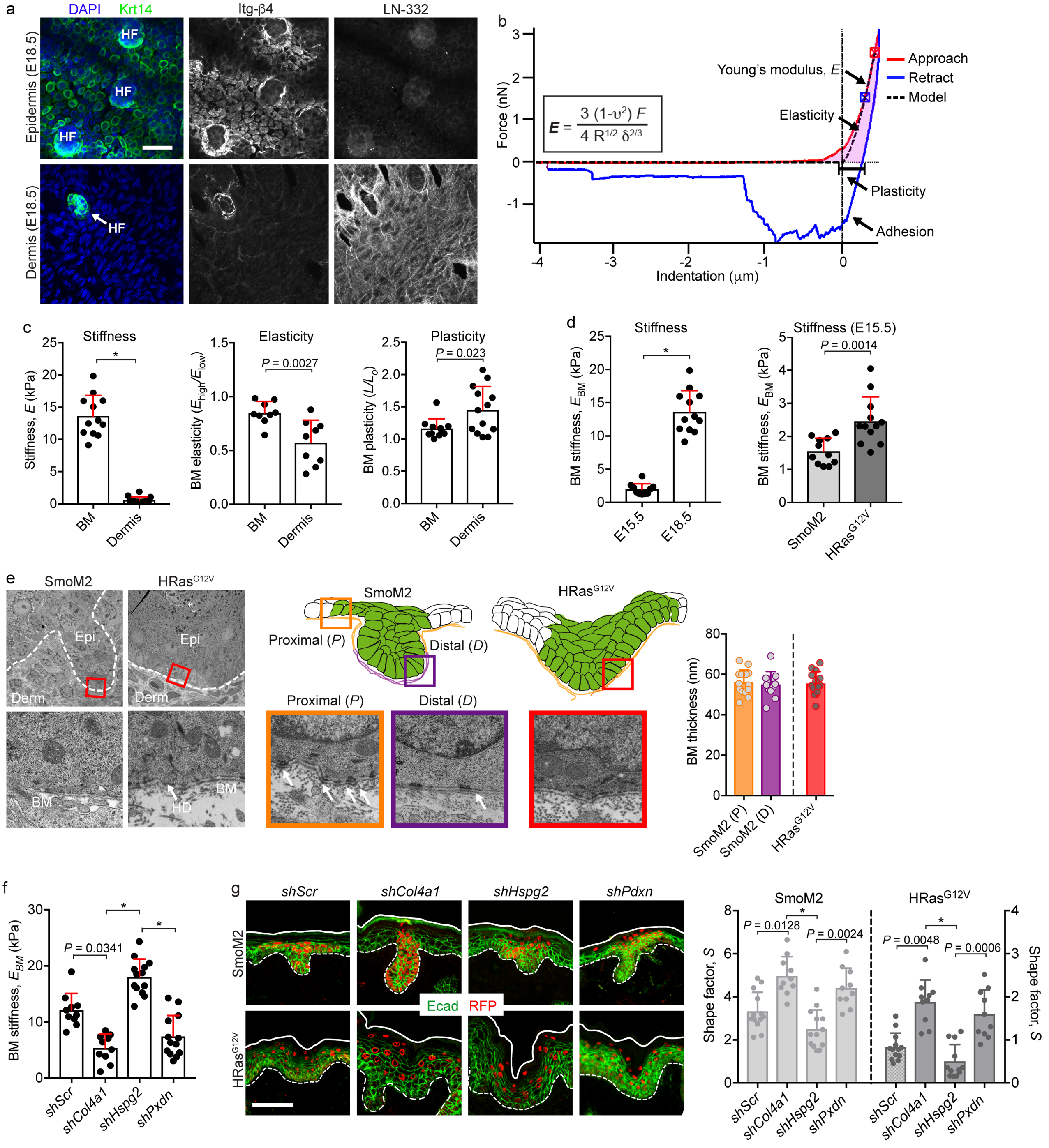


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Role of interfacial actomyosin tension in a monolayer and multilayered epithelium. **a**, Left, example time-lapse kymographs showing junctional laser ablation. Plasma membranes are marked by membrane-Tomato and membrane-GFP (mT and mG) in SmoM2;mTmG and HRas^{G12V};mTmG mice, with mutant cells (M) in green, wild-type (WT) cells in blue, and M–WT/WT–WT/M–M interfaces labelled. Laser cut sites are marked with pink dashed lines. Centre and right, the displacement of neighbouring tricellular junctions was quantified over time to yield retraction velocity curves. Initial retraction velocity values are shown for WT ($n = 13$), SmoM2, (M–WT, $n = 9$; M–M, $n = 17$; WT–WT, $n = 8$) and HRas^{G12V} (M–WT, $n = 10$; M–M, $n = 16$; WT–WT, $n = 17$; one-way ANOVA with Tukey's multiple comparisons test) from four to five embryos from two litters for each condition.
b, Immunofluorescence staining of F-actin (using phalloidin) and phospho-S19-myosin-II (p-MyoII) in SmoM2 and HRas^{G12V} lesions in sagittal sections (left) and planar whole-mount (right) views at E15.5. The intensity of staining is shown in heatmap values. p-MyoII polarization was measured in single basal cells (along the apicobasal axis, sagittal view; WT, $n = 15$; SmoM2, $n = 18$; HRas^{G12V}, $n = 16$; one-way ANOVA with Tukey's multiple comparisons test) and in whole clones (M–WT versus M–M interface, planar view; WT, $n = 12$; SmoM2, $n = 12$; HRas^{G12V}, $n = 16$). Note that although p-MyoII is enriched basally, this polarization does not change between WT epidermal progenitors and oncogenic basal cells.
c, Cell shapes analysed from E15.5 SmoM2 mutant clones. Cell area and anisotropy (defined as the ratio of major and minor cell axes) were analysed from whole-mount confocal images. Cells were automatically segmented on the basis of cortical E-cadherin staining. Note the increased anisotropy in M and WT cells at the clone border and the diminished cell area at the clone centre. **d**, The monolayer model epithelium. A single cell is transformed (green) and then undergoes cycles of division to induce tissue growth and deformation (see Supplementary Note 1). Interfacial tensions were varied in magnitude and orientation from basally to apically polarized, resulting in evaginating or invaginating lesions, respectively. **e**, Explant cultures treated with the actomyosin inhibitor Y-27632. SmoM2 oncogenic skin explants were treated with Y-27632 or vehicle control (DMSO) for 24 h before preparing the tissue for microscopic analysis ($n = 11$ lesions from three explants each; two-tailed unpaired t-test). All bar graphs show means + s.d. Scale bars, 50 µm. $^*P < 0.0001$.

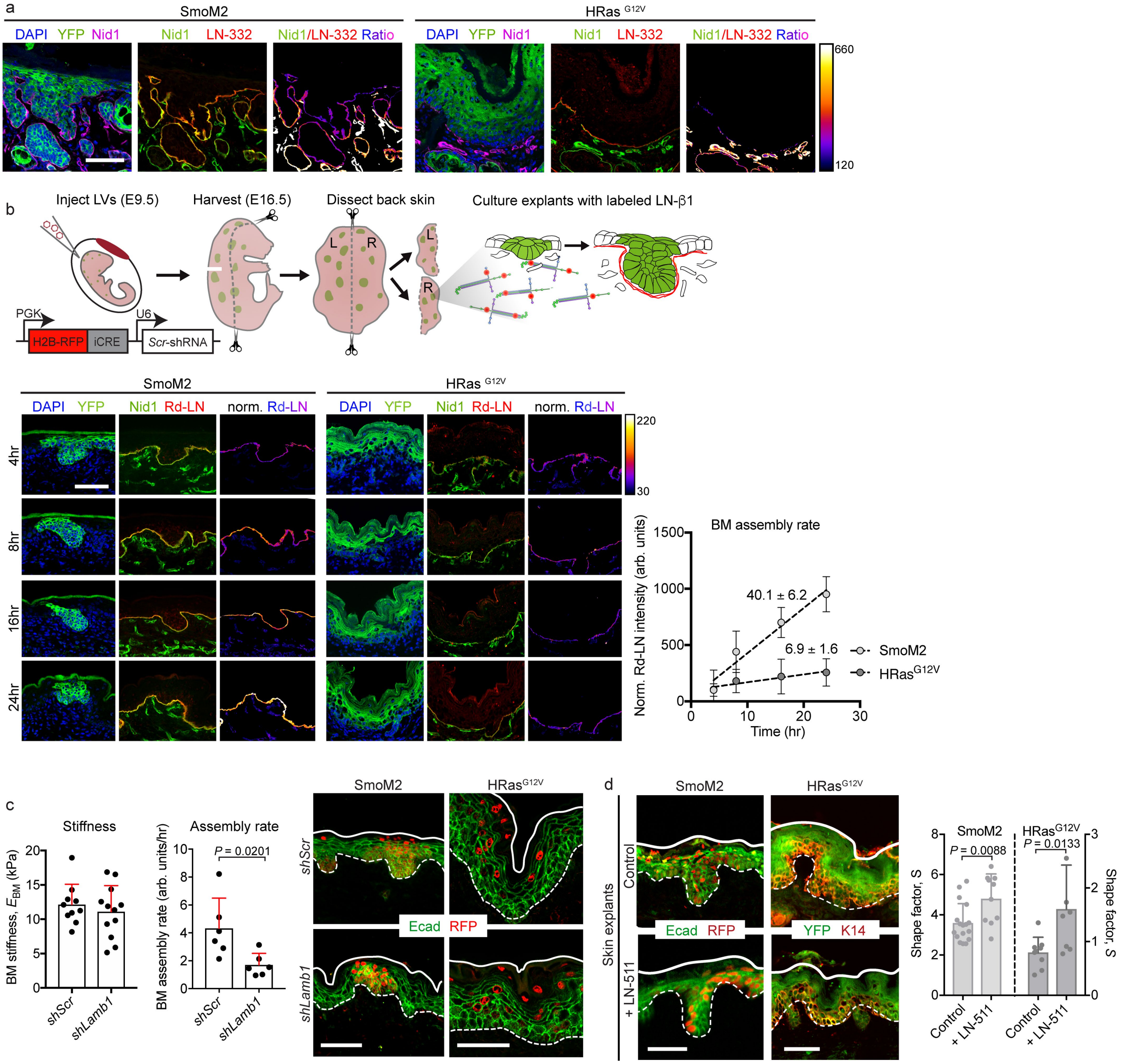


Extended Data Fig. 4 | FACS sorting strategy and RNA-sequencing analysis of SmoM2 and HRas^{G12V} tumours. a, FACS strategy for isolating fluorescently marked basement membrane (BM)-associated (α6 integrin^{hi} RFP^{+}) basal progenitors from WT or oncogenic skins of E15.5 embryos. **b**, Principal component analysis (PCA) plots of $n = 3$ independent replicates of E15.5 oncogenic and WT basal progenitors reveals clustering of each replicate but distinct clustering across genetic lineages. **c**, Venn diagram showing genes upregulated or downregulated, comparing SmoM2 or HRas^{G12V} mutants to WT basal progenitors. The overlap shows that 167 genes (6% of these differentially expressed genes) were coordinately upregulated in SmoM2 and downregulated in HRas^{G12V} mutants.

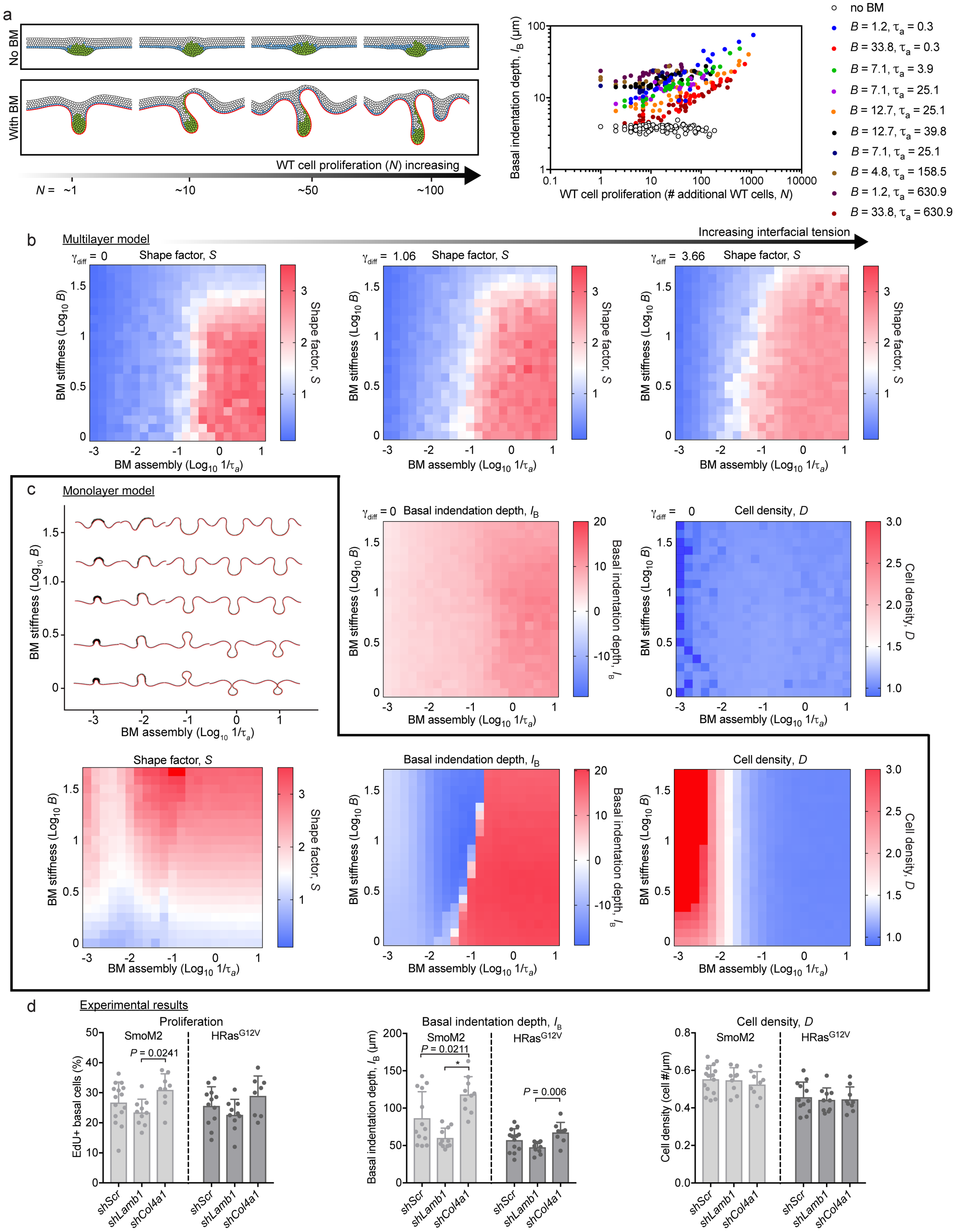


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Extended characterization of the mechanical properties of basement membrane. a, Immunofluorescence images of the epidermal and dermal interfaces following EDTA-induced skin separation. Integrins (marked here by Itg-β4) on the progenitors' basal surface delineate the underside of Krt14$^+$ epidermis, while the basement membrane (marked here by LN-332) delineates the dermal surface of the tissue. To prepare samples for AFM measurements of basement-membrane stiffness, E_{BM}, lentivirus-infected skin is harvested, and the epidermis is separated from dermis using EDTA treatment, leaving the basement membrane exposed on the dermal surface. HF, hair follicles. **b**, Example of AFM data analysis: a force-displacement curve for basement membrane, showing approach (red) and retraction (blue) curves, as well as the contact point (crosshairs). The equation for the Hertz model used to calculate E_{BM} and its corresponding curve fit (black dotted line) is also shown. **c**, Stiffness, elasticity and plasticity of the basement membrane and dermis. The elasticity metric is defined as $E_{\mathrm{low}}/E_{\mathrm{high}}$, where a value of one represents absolute linear elasticity and values less than are increasingly nonlinear. The plasticity metric (L/L_0) is defined as the difference in indentation lengths at zero force values between the approach and retraction curves. Measurements are from average force maps (basement membrane, $n = 12$; dermis, $n = 14$; two-tailed unpaired *t*-test) from four WT embryos. **d**, Stiffening of the basement membrane during epidermis and tumour development. AFM measurements are shown for WT basement membrane (left; E15.5, $n = 10$; E18.5, $n = 12$; Mann–Whitney *U*-test) and oncogenic lesions at E15.5 (right; SmoM2, $n = 11$; HRas$^{\mathrm{G12V}}$, $n = 12$; two-tailed unpaired *t*-test) and E18.5. The data for basement-membrane stiffness at E18.5 are the same as in **c** for purposes of comparison. **e**, Ultrastructural measurements. TEM images of the indicated regions show basement membrane (BM), dermis (Derm), epidermis (Epi) and hemidesmosomes (HD). Basement-membrane thickness is also quantified (SmoM2 (*P*), $n = 14$; SmoM2 (*D*), $n = 12$; HRas$^{\mathrm{G12V}}$, $n = 14$; one-way ANOVA with Tukey's multiple comparisons). **f**, E_{BM} measurements of shRNA-transduced and EDTA-treated skins (*shScr*, $n = 9$; *shCol4a1*, $n = 9$; *shHspg2*, $n = 13$; *shPxdn*, $n = 13$; Kruskal–Wallis test with Dunn's multiple comparisons) from three embryos each. **g**, Representative immunofluorescence images of oncogenic skins from SmoM2 or HRas$^{\mathrm{G12V}}$ embryos transduced with LV–Cre harbouring either *Scr*, *Col4a1*, *Hspg2* or *Pdxn* shRNAs. *S* values are quantified (SmoM2: *shScr*, $n = 13$; *shCol4a1*, $n = 10$; *shHspg2*, $n = 10$; *shPxdn*, $n = 10$. HRas$^{\mathrm{G12V}}$: *shScr*, $n = 13$; *shCol4a1*, $n = 11$; *shHspg2*, $n = 12$; *shPxdn*, $n = 11$; Kruskal–Wallis test with Dunn's multiple comparisons) from four embryos, two litters for each condition. All bar graphs show means + s.d. Scale bars, 50 µm. $^*P < 0.0001$.

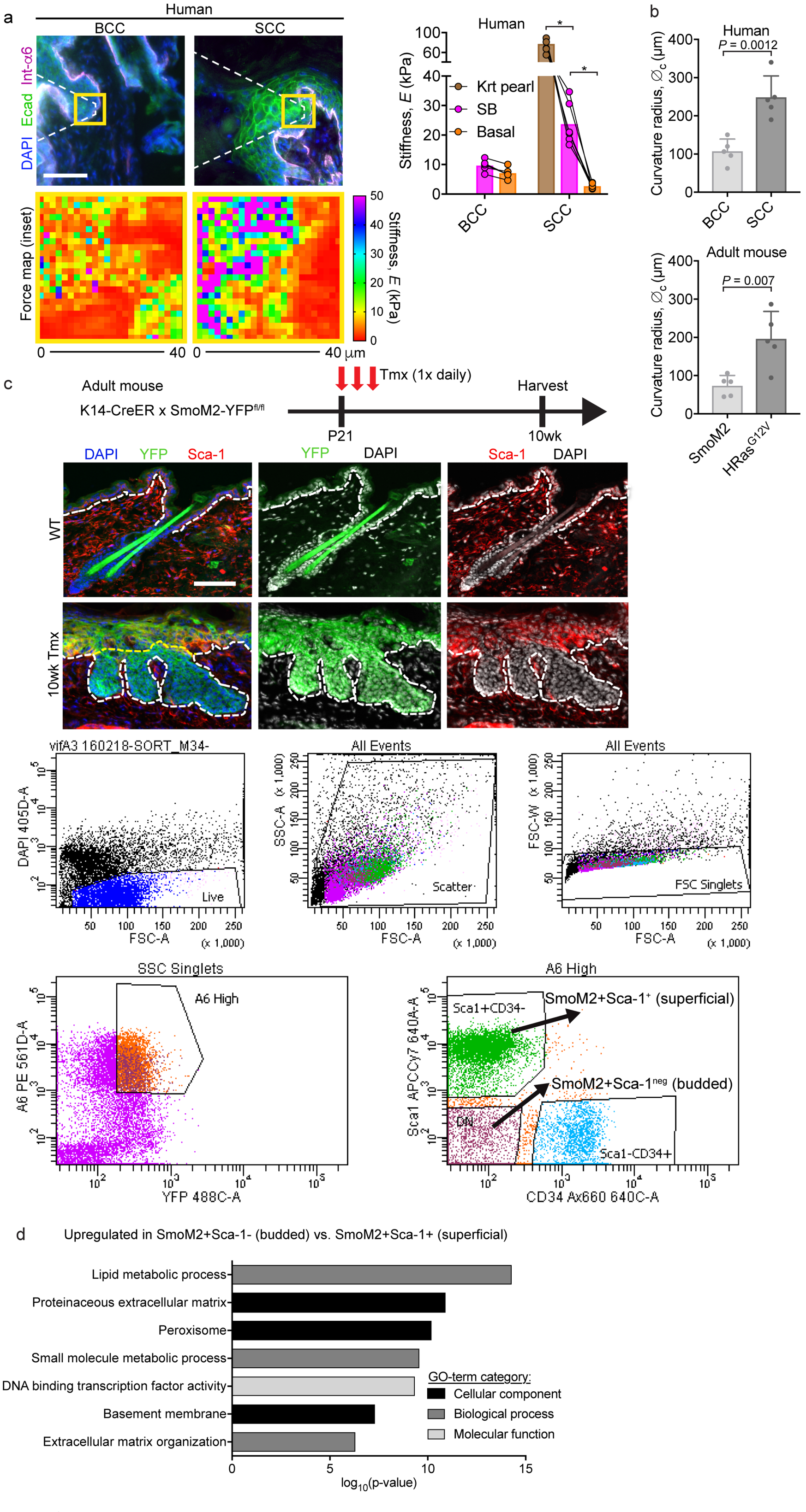


Extended Data Fig. 6 | Extended characterization of the effects of basement-membrane assembly on tumour architectures.
a, Immunofluorescence of oncogenic skin sections. Note the enriched expression of nidogen (Nid1, a component of nascent basement membranes) at the leading edge of SmoM2 lesions, compared with the expression of LN-332 (a component of mature basement membranes), as shown by the ratiometric intensity heatmap. **b**, Our experimental approach for assaying the assembly rate of basement membranes, and experimental results. Oncogenic skin explants were harvested, and basement-membrane assembly rates were measured by exogenous pulse-labelling with fluorescent LN-β1 (Rd-LN), measuring its incorporation into native membranes over time. Example images are shown at the bottom left, with the intensity ratio of labelled LN-β1 and endogenous Nid1 immunostaining shown in heatmap. Bottom right, quantification of the LN-β1/Nid1 intensity ratio over time, with the slope (*m*) of each linear fit indicated. **c**, Changes in basement-membrane stiffness and assembly resulting from *Lamb1* knockdown. Left, E_{BM} measurements from shRNA-transduced WT embryos (*shScr*, $n = 10$; *shLamb1*, $n = 12$; independent regions from three embryos each; mean + s.d.; two-tailed unpaired *t*-test). Centre, assembly rates were measured as in **b** (*shScr*, $n = 6$; *shLamb1*, $n = 6$; Mann–Whitney *U*-test). Right, representative immunofluorescence images. **d**, Gain-of-function effects of LN-α5β1γ1 (LN-511) on tissue architecture. E16.5 oncogenic skin explants were cultured for 24 h with an excess ($100\ \mu g\ ml^{-1}$) of soluble LN-511 or vehicle control. Left, immunofluorescence staining for E-cadherin and RFP (SmoM2) or YFP and K14 (HRas^{G12V}). Transduced cells are RFP^{+} and YFP^{+}, respectively. Right, lesional *S* measurements for LN-511-treated oncogenic explants (SmoM2: control, $n = 17$; + LN-511, $n = 10$. HRas^{G12V}: control, $n = 9$; + LN-511, $n = 8$; two-tailed unpaired *t*-test) from four embryos. All bar graphs show means + s.d. Scale bars, 50 µm.



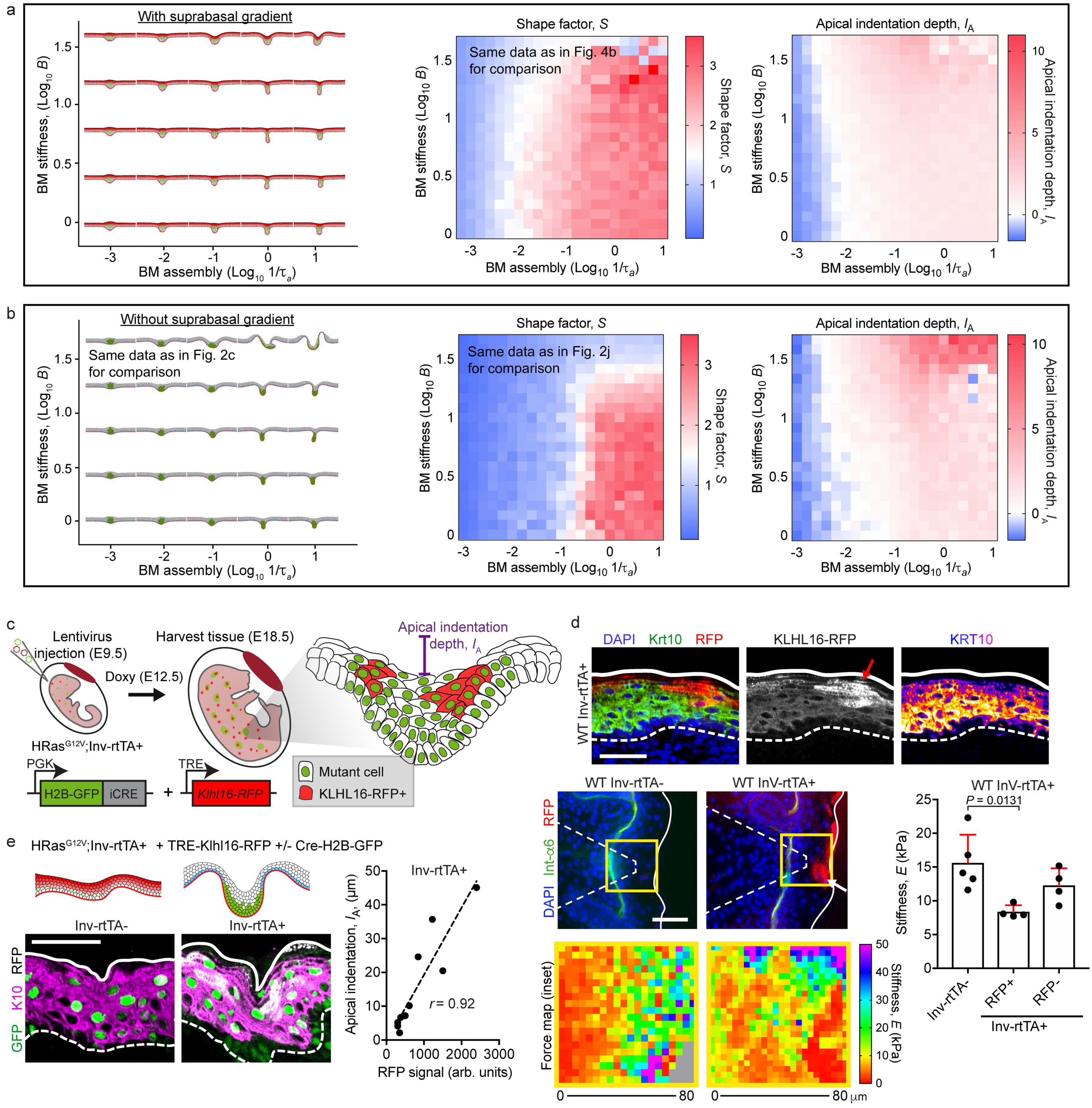
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Extended characterization of the effects of biophysical properties of the basement membrane on tumour architecture. a, Left, a multilayer simulation was constructed to include WT cell proliferation in the absence or presence of the basement membrane. Simulated tissues with varying final numbers of WT cells (*N*) are shown. Right, basal indentation depth (I_B) is quantified for varying values of basement-membrane stiffness (*B*) and assembly rate ($1/\tau_a$). The extent of tissue deformations increases with increasing *N*, and this trend is globally conserved across multiple orders of magnitude and biophysical properties of basement membrane. **b**, The effect of interfacial tension on multilayer epithelia in the presence of basement membrane. Phase diagrams of *S* are shown are for $\gamma_{\text{diff}} = 0, 1.06$ and 3.66. The effect is to gradually increase *S* values for an increasingly broad range of biophysical properties of the basement membrane. *S*, I_B and cell density (*D*) phase diagrams from the multilayer model are shown for comparison with the monolayer simulations in **c**. **c**, The monolayer model, with mechanical properties of the basement membrane adjusted directly beneath transformed cells (see Supplementary Note 1). Shown (top) are phase diagrams for simulated tissue shapes predicted by the monolayer model as the stiffness and assembly rates of basement membrane are varied over the full parameter space, as well as (bottom) *S*, I_B, and *D* simulations predicted by the monolayer model. **d**, Experimental measurements of cell density and proliferation in E18.5 WT, SmoM2 and HRas^{G12V} skins and in oncogenic lesions transduced by *shScr*, *shLamb1* or *shCol4a1*. Data are from four embryos, two litters for each condition (Kruskal–Wallis test with Dunn's multiple comparisons). All bar graphs show means + s.d.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Measuring tissue mechanics, architecture and gene expression in human and adult mouse BCCs and SCCs. **a**, AFM measurements of the stiffness of tumour compartments in human BCCs and SCCs. Left, immunofluorescence images of BCCs and SCCs, with force maps of the boxed areas shown below each image. Right, graph showing stiffness values for basal, keratin pearl (Krt Pearl) and suprabasal (SB) regions from BCCs ($n = 6$ regions from four tumours) and SCCs ($n = 6$ regions from three tumours). Paired measurements are compared between tumour type and tumour compartment (two-way ANOVA with Tukey's multiple comparisons test). **b**, Curvature radius ($\varnothing_c$) values for human and mouse BCCs and SCCs ($n = 5$ tumours for each group; two-tailed unpaired t-test). **c**, Strategy for inducing SmoM2 in mice at postnatal day (P)21 and then harvesting 10 weeks later for FACS isolation and transcriptional profiling of $\alpha6^{hi}$ YFP^{+} SmoM2^{+} basal progenitors from budded (Sca1^{-}) and superficial (non-budded, Sca1^{+}) tissue. **d**, GO terms for mRNAs upregulated in budded versus superficial BCC progenitors. Note that the ECM and basement-membrane categories are particularly enriched in budded progenitors. All bar graphs show means + s.d. Scale bars, 50 μm. *$P < 0.0001$.



Extended Data Fig. 9 | Effects of a suprabasal stiffness gradient on tissue architectures. **a**, Right, phase diagrams of S and I_A for the simulated tissue shapes shown at the left with the multilayer cell stiffness gradient for varying basement-membrane stiffness (B) and assembly rates ($1/\tau_a$). For comparison, we show the same phase diagram for S as in Fig. 4b. **b**, Right, phase diagrams of S and I_A for the simulated tissue shapes shown at the left without the multilayer cell stiffness gradient for varying B and $1/\tau_a$. For comparison, we show the same tissue shapes and phase diagram for S as in Fig. 2c, j, respectively. **c**, Manipulation of suprabasal cell stiffness with the ubiquitin ligase KLHL16. TRE–KLHL16–RFP was induced in suprabasal cells after treating Inv–rtTA mice with doxycycline at E9.5. **d**, Top, immunofluorescence staining shows a decrease in K10 intensity that overlaps with the RFP signal (arrow). Bottom left, AFM force maps from WT Inv–rtTA$^-$ and Inv–rtTA$^+$ embryos at E18.5. Note the decreased stiffness that correlates with the RFP signal in Inv–rtTA$^+$ embryos (arrow). Right, force maps were quantified and compared between RFP$^-$ and RFP$^+$ regions of Inv–rtTA$^+$ embryos and Inv–rtTA$^-$ embryos (mean + s.d., one-way ANOVA with Tukey's multiple comparisons test). **e**, TRE–KLHL16–RFP was induced in suprabasal cells in the HRas^{G12V} background using LV–Cre–H2B–GFP. Tissues harvested at E18.5 were analysed for I_A, which correlated linearly with the extent of the TRE–KLHL16–RFP signal (r, Pearson's correlation coefficient; $n = 11$ regions from four embryos). Scale bars, 50 μm.

nature research

Corresponding author(s): Elaine Fuchs

Last updated by author(s): Jun 10, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Images were acquired using ZEN Blue (Carl Zeiss, v2.3), iQ3 (Andor, v3.6.2) and Asylum Research v14 (IgorPro 6.2.1) FACS data was collected using FACSDiva (BD Biosciences, v8.0.3) software. Simulations were run using C/C++ based on standard C libraries and GNU Scientific Library (GSL, v2.6).
Data analysis	Images were analysed using Fiji (NIH, 2.0.0-rc-69/1.52p), CellProfiler (v3.1.8) and Imaris (Oxford Instruments, v8.3.1). Sequencing data were analysed using STAR (v2.6.2a) and DEseq2 (v1.24.0). Data were compiled and statistical tests were performed using Excel (Microsoft, v14.5.7), Prism (Graphpad, v7), R (version 3.4.4) and RStudio (Version 1.1.442). Figures were assembled using Adobe Illustrator CS6 (v16.0.0).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All RNA-sequencing data are in the processes of being deposited into the Gene Expression Omnibus and accession codes will be made available before publication. All other data in the manuscript, supplementary materials, and source data are available from the authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes. Sample sizes were determined based on previous publications on similar experiments (Nature volume 501, pages185–190, 2013).
Data exclusions	No data were excluded.
Replication	Every experiment was performed on at least 2 independent litters. All attempts at replication were successful.
Randomization	Samples (mouse embryos) were allocated randomly into experimental groups.
Blinding	Investigators were not blinded to experimental groups/genotypes due to the nature of embryo recovery.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☐	☒	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☐	☒	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	Antibodies used were as follows: rat anti-RFP (1:1000, Chromotek, 15F8, lot# 60706002AB), rabbit anti-RFP (1:1000, MBL, PM005, lot# 044), chicken anti-GFP (1:2000, Abcam, ab13970, lot# GR3190550-6), goat anti-P-Cadherin (1:500, R&D, AF761), rabbit anti-E-cadherin (1:500, Cell Signaling Technology, 1:500, 9835, lot# GR3190550-6), rat anti-E-cadherin (1:200, M. Takeichi), guinea pig anti-keratin14 (1:500, Fuchs laboratory), rabbit anti-keratin10 (1:1000, Covance, poly19054, lot# D15LF02452), rabbit anti-collagen type IV (1:500, Abcam, ab6586), rat anti-nidogen (1:200, Santa Cruz Biotechnology, ELM1), rat anti-Laminin-β1 (1:100, Abcam, LT3), rabbit anti-Laminin-α5 (1:500, J. Miner, Washington University in St. Louis), rabbit anti-Laminin-332 (1:500, P. Marinkovich, Stanford University), mouse anti-phospho-S22-Myosin light chain 2 (1:100, Cell Signaling Technology, 3675), mouse anti-vimentin (1:200, Dako, 3B4), rat anti-Sca-1 (1:200, Becton Dickinson, D7), rat biotinylated anti-CD45 (1:200, BD Bioscience, 553078, lot# 6294639), rat biotinylated anti-CD140a (1:200, Biolegend, 135910, lot# B236869), rat biotinylated anti-CD31 (1:200, Biolegend, 102504, lot# B208712), rat biotinylated anti-cKit (1:200, Biolegend, cat#105804, lot#B208999), rat PECy7-conjugated-anti alpha6/CD49f (1:1000, Biolegend, 313621, lot# B211003). All secondary antibodies used were raised in a donkey host and were conjugated to one of AlexaFluor488, AlexaFluor546 or AlexaFluor647 (1:500, Life Technologies).
Validation	All antibodies are commercially available and validated by the manufacturer except for rabbit anti-keratin14 (Fuchs lab), rabbit anti-Laminin-α5 (J. Miner) and rabbit anti-Laminin-332 (P. Marinkovich), which were used at the stated concentrations. The subcellular localization of all the proteins analyzed in this study has been previously reported. This was used to validate the specificity of the antibody.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Animals were fed regular rodent chow with ad libitum access to food and water. Euthanasia was by CO2 asphyxiation. The following

Laboratory animals	previously generated mouse (Mus musculus) lines were used in this study: Rosa26-SmoM2-YFPfl/fl, FrHRas-G12Vfl/fl, Rosa26-EYFPfl/fl, Rosa26-mTmG, Krt14-rtTA, hIVL-rtTA. C57Bl6J/CD1 mixed background strains were used. Embryos were analyzed from E12.5 to E18.5, and adult mice were collected at P21 and 3-6mo for tumor analysis. Both sexes were used for adult studies, and sex was not determined in embryos.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All procedures were approved by the Rockefeller University IACUC and performed within an AAALAC-certified animal facility.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	De-identified OCT-embedded fresh, non-fixed tissue blocks of SCCs, BCCs or healthy skin from male and females (age range 49-82) that underwent Mohs micrographic surgery were used.
Recruitment	De-identified tissue blocks were obtained from the Department of Dermatology, Weill-Cornell Medical College (New York, US). This study did not involve the recruitment of new patients, and no self-selection bias is acknowledged.
Ethics oversight	We have complied with all relevant ethical regulations, including obtaining informed consent from patients by Weill-Cornell and the Rockefeller University IRB approved the use of de-identified human samples (EFU-0529).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Dorsal skins were surgically removed from E15.5 embryos and incubated for 30min at 37C in 1:1 trypsin:versene. Cells were filtered through a 40um cell strainer (VWR) to generate a single cell suspension.
Instrument	BD FACSAriaII, 70um nozzle
Software	BD FACSDIva Software
Cell population abundance	Lin-;LV-Cre+ cell populations were about 5% of the total population at E15.5.
Gating strategy	Cells were gated on DAPI-negative (for live/dead) and FSC/SSC singlets. Lineage-negative cells were identified by staining with biotinylated primary antibodies against CD45 (to exclude immune cells), CD117/cKit (to exclude the melanoblast lineage), CD31 (to exclude endothelial lineage), and CD140a (to exclude fibroblasts). All lin- cells were labelled with a streptavidin-PECy7 conjugated secondary antibody. CD49f/alpha6-high cells (a marker of the basal epidermis) that double labelled with RFP (as a marker of LV-Cre transduction) were isolated to identify WT, SmoM2+ and HRas+ LV-Cre transfected cells (in control, Rosa26-SmoM2-YFPfl/fl and FrHRas-G12Vfl/fl;Rosa26-EYFPfl/+ background genotypes, respectively). Both sexes were used as sex was not determined in embryos). Each sample comprised single cells from at least 3 embryos per genotype.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

TRIM37 controls cancer-specific vulnerability to PLK4 inhibition

https://doi.org/10.1038/s41586-020-2710-1

Received: 11 June 2019

Accepted: 9 June 2020

Franz Meitinger[1]✉, Midori Ohta[1], Kian-Yong Lee[1], Sadanori Watanabe[1], Robert L. Davis[2], John V. Anzola[2], Ruth Kabeche[1], David A. Jenkins[2], Andrew K. Shiau[2,3], Arshad Desai[1,3,4,5]✉ & Karen Oegema[1,3,4,5]✉

Centrosomes catalyse the formation of microtubules needed to assemble the mitotic spindle apparatus[1]. Centrosomes themselves duplicate once per cell cycle, in a process that is controlled by the serine/threonine protein kinase PLK4 (refs. [2,3]). When PLK4 is chemically inhibited, cell division proceeds without centrosome duplication, generating centrosome-less cells that exhibit delayed, acentrosomal spindle assembly[4]. Whether PLK4 inhibitors can be leveraged as a treatment for cancer is not yet clear. Here we show that acentrosomal spindle assembly following PLK4 inhibition depends on levels of the centrosomal ubiquitin ligase TRIM37. Low TRIM37 levels accelerate acentrosomal spindle assembly and improve proliferation following PLK4 inhibition, whereas high TRIM37 levels inhibit acentrosomal spindle assembly, leading to mitotic failure and cessation of proliferation. The Chr17q region containing the *TRIM37* gene is frequently amplified in neuroblastoma and in breast cancer[5–8], rendering these cancer types highly sensitive to PLK4 inhibition. We find that inactivating TRIM37 improves acentrosomal mitosis because TRIM37 prevents PLK4 from self-assembling into centrosome-independent condensates that serve as ectopic microtubule-organizing centres. By contrast, elevated TRIM37 expression inhibits acentrosomal spindle assembly through a distinct mechanism that involves degradation of the centrosomal component CEP192. Thus, TRIM37 is an essential determinant of mitotic vulnerability to PLK4 inhibition. Linkage of *TRIM37* to prevalent cancer-associated genomic changes—including 17q gain in neuroblastoma and 17q23 amplification in breast cancer—may offer an opportunity to use PLK4 inhibition to trigger selective mitotic failure and provide new avenues to treatments for these cancers.

Cells entering mitosis have two centrosomes that catalyse the generation of microtubules for assembly of the mitotic spindle[1]. Each centrosome has a centriole at its core that recruits a proteinaceous matrix called the pericentriolar material that nucleates and anchors microtubules[9]. Centrioles duplicate in a cell-cycle-coupled process that is controlled by the Polo-family kinase PLK4 (refs. [2,3]). To explore centrosome biology and the potential of inhibiting PLK4 as a treatment for cancer, the selective and cellularly active PLK4 inhibitor centrinone was previously developed[4,10]. In the presence of centrinone, continued cell division without centriole duplication generates centrosome-less cells[4]. Cells lacking centrosomes remain capable of forming a bipolar spindle; however, spindle assembly and chromosome alignment are delayed and error-prone[4,11–14]. Chromosome segregation fails in roughly 10% of non-transformed human RPE1 cells treated with centrinone, leading to eventual growth arrest[13].

TRIM37 controls cell response to centrinone

In a genome-wide screen for genes whose inactivation enables sustained proliferation of centrinone-treated RPE1 cells, the ubiquitin ligase TRIM37 was previously identified[13]. We find now that loss of *TRIM37* does not alter the duration of mitosis for cells with centrosomes (treated with vehicle, dimethylsulfoxide (DMSO)), but does rescue the delayed spindle assembly and chromosome-segregation failure seen in cells that lack centrosomes (treated with centrinone)[13] (Fig. 1a, b, Extended Data Fig. 1a–e and Supplementary Video S1). To determine whether elevating TRIM37 levels has the opposite effect, we conditionally overexpressed TRIM37 (Extended Data Fig. 1a–c). A roughly fourfold increase in TRIM37 levels did not affect mitotic timing in cells with centrosomes, but significantly increased mitotic duration and chromosome-segregation failure in centrinone-treated cells (Fig. 1a, b, Extended Data Fig. 1d, e and Supplementary Video S1; $P < 0.0001$). Analysis of four additional cell clones with varying increases in TRIM37 levels indicated that the magnitude of the mitotic defects in centrinone-treated cells was proportional to the amount of TRIM37 (Extended Data Fig. 1c, f). Thus, the extent of mitotic challenge imposed by centrosome loss following PLK4 inhibition depends on TRIM37 in a bidirectional fashion: loss of TRIM37 improves outcomes, whereas increases in TRIM37 substantially worsens them.

---

[1]Ludwig Institute for Cancer Research, La Jolla, CA, USA. [2]Small Molecule Discovery Program, Ludwig Institute for Cancer Research, La Jolla, CA, USA. [3]Section of Cell and Developmental Biology, Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA. [4]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA. [5]These authors contributed equally: Arshad Desai, Karen Oegema. ✉e-mail: fmeitinger@ucsd.edu; abdesai@ucsd.edu; koegema@ucsd.edu

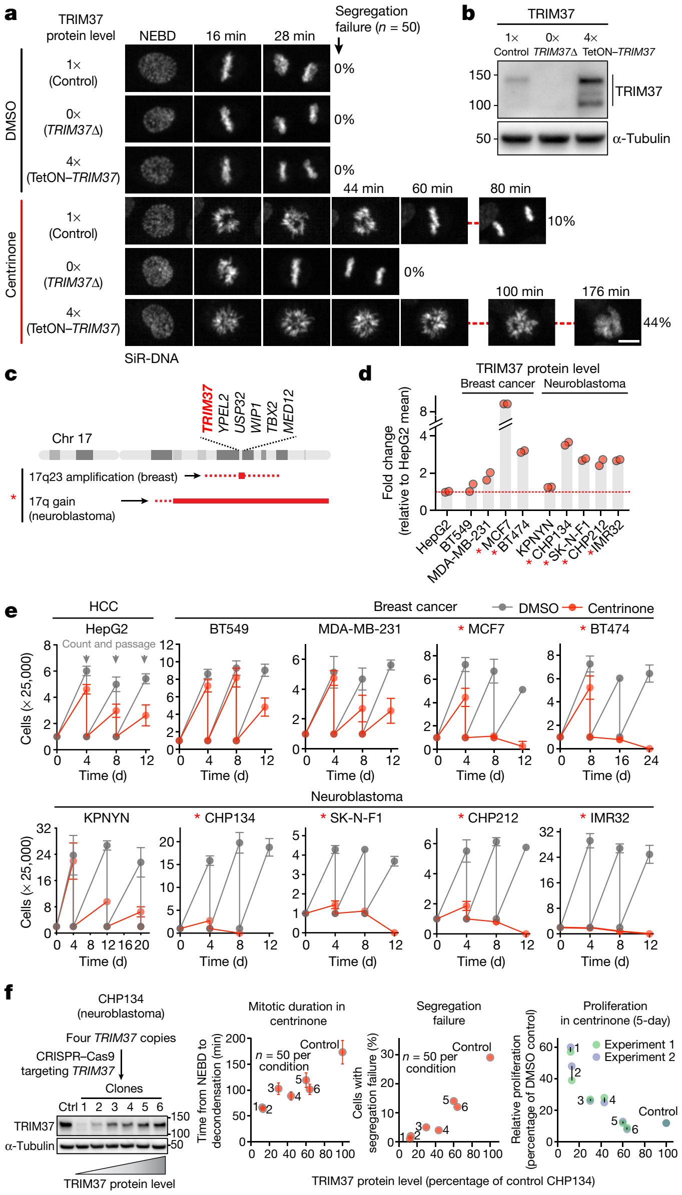


Fig. 1 | TRIM37 levels determine mitotic outcomes and cancer-specific sensitivity to PLK4 inhibition. **a**, Still images from time lapse sequences showing chromosomes in RPE1 cells with normal (1×), no (0×, *TRIM37Δ*) or fourfold increased (4×) TRIM37 protein levels after treatment with DMSO or centrinone. Scale bar, 10 µm. Rates of chromosome-segregation failure are also indicated. Overexpression of TRIM37 was achieved using an inducible transgene (TetON–*TRIM37*). **b**, Immunoblot showing TRIM37 levels in the three lines analysed; α-tubulin serves as a loading control. **c**, Diagram of the chromosome 17q region containing *TRIM37* that is amplified in specific cancer contexts. The amplified region is indicated with a solid red line and adjacent non-amplified regions with dashed lines. **d**, TRIM37 protein levels, measured by semiquantitative immunoblotting, in the indicated breast cancer and neuroblastoma cell lines. Asterisks mark cell lines with *TRIM37* amplification. **e**, Passaging-based proliferation analysis for the indicated cell lines treated with DMSO or centrinone. **f**, Left, immunoblot of CHP134 clones in which CRISPR–Cas9-based inactivation of one or more of the four *TRIM37* gene copies was used to vary TRIM37 protein levels. α-Tubulin serves as a loading control. Right, graphs plotting mitotic duration, frequency of chromosome-segregation failure, and proliferation in centrinone as a function of TRIM37 protein level in the engineered CHP134 cell lines. Error bars for mitotic duration represent 95% confidence intervals. For gel source data, see Supplementary Fig. 1.

TRIM37 elevation in specific cancers

The *TRIM37* locus is found at the border of 17q22 and 17q23—a chromosomal region that is amplified in a number of cancers, most prominently in around 50–60% of neuroblastomas and roughly 10% of breast cancers[5–8] (Fig. 1c). Consistent with the prevalence of 17q amplification in neuroblastomas[6], levels of *TRIM37* messenger RNA are significantly higher in neuroblastomas compared with other paediatric cancers[15] (Extended Data Fig. 1g; $P < 0.0001$). As expected from the tumour expression data, cell lines derived from neuroblastomas and a subset of breast cancers also exhibit high *TRIM37* expression[16] (Extended Data Fig. 1h, i). To assess whether elevated TRIM37 expression in cancers confers enhanced sensitivity to PLK4 inhibition, we analysed two breast cancer (BT474 and MCF7) and four neuroblastoma (CHP134, SK-N-F1, CHP212 and IMR32) cell lines with amplification of *TRIM37*; four cancer cell lines that lack *TRIM37* amplification—derived from neuroblastoma (KPNYN), breast cancer (BT549 and MDA-MB-231) and hepatic cancer (HepG2)—served as controls (Extended Data Fig. 1j). Immunoblotting confirmed increases in TRIM37 protein levels in cell lines with *TRIM37* amplification (Fig. 1d and Extended Data Fig. 2a–c). Passaging-based proliferation analysis revealed that non-amplified cancer cell lines behaved similarly to the more than 20 previously characterized cancer cell lines[4], in that they continued to proliferate in centrinone, albeit at a reduced rate owing to increased mitotic errors (Fig. 1e and Extended Data Fig. 2b) (centrosome depletion was confirmed in these cell lines[4]; Extended Data Fig. 2d). By contrast, the six cancer cell lines with elevated TRIM37 levels failed to proliferate in centrinone, suggesting synthetic lethality with PLK4 inhibition (Fig. 1e).

To address the causal relationship between cancer-specific elevation of TRIM37 and sensitivity to PLK4 inhibition, we used CHP134 neuroblastoma cells, which exhibit high sensitivity to centrinone (Extended Data Fig. 2e–g and Supplementary Video S2). As these cells have four genes encoding *TRIM37*, variable targeting of the distinct gene copies enabled us to generate a six-clone ‘allelic series’, with TRIM37 protein levels ranging from roughly 10% to 70% of those in parental CHP134 cells (Fig. 1f). Live imaging revealed a notable correlation between the amount of TRIM37 and the severity of the mitotic defects following PLK4 inhibition (Fig. 1f and Extended Data Fig. 2h). Moreover, as these data would predict, proliferation in centrinone correlated inversely with TRIM37 protein levels (Fig. 1f and Extended Data Fig. 2h). Thus, in the context of a *TRIM37*-amplified cancer cell line, TRIM37 levels dictate sensitivity to PLK4 inhibition.

TRIM37 prevents PLK4 condensation

We next investigated why reducing TRIM37 levels improves acentrosomal mitosis, whereas increasing TRIM37 levels renders it prone to fail. Surprisingly, the results indicate that the effects of decreasing versus increasing TRIM37 protein levels are mechanistically distinct. In RPE1 and CHP134 cells with reduced levels of TRIM37, PLK4 was found both at centrosomes and, frequently, in a single large condensate distinct from the centrosome[13] (Fig. 2a, d and Extended Data Fig. 3a); condensate formation was not a consequence of increased PLK4 abundance (Fig. 2b and Extended Data Fig. 3b). Of 12 tested centrosomal proteins, including 2 (CEP192 and CEP152) that interact with PLK4 (ref. [2]), only PLK4 was found in the ectopic condensate (Extended Data Fig. 3c, d). The PLK4-containing condensate in *TRIM37Δ* cells did not nucleate microtubules during interphase (data not shown); however, as cells progressed into mitosis, roughly 25% of condensates acquired additional centrosomal components and nucleated microtubules (Extended Data Fig. 3e). These ectopic microtubule-generating centres clustered with one of the two centrosome-based spindle poles, resulting in bipolar division (data not shown). In centrinone-treated *TRIM37Δ* cells, instead of a single large condensate, PLK4 was present in an array of

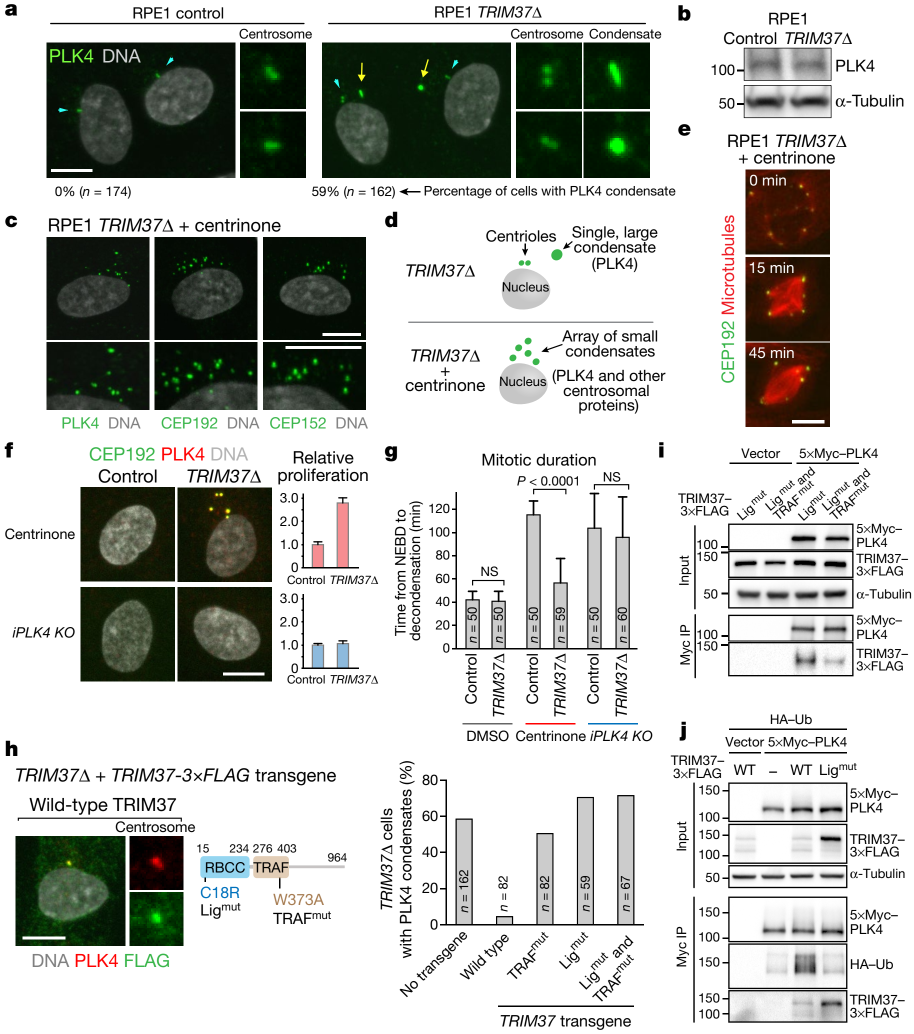


Fig. 2 | TRIM37 prevents the formation of ectopic microtubule-organizing centres based on PLK4 condensates. a, Immunofluorescence images showing the localization of PLK4 in interphase RPE1 cells. Cyan arrowheads show PLK4 localization at centrioles; yellow arrows show PLK4 localization to single ectopic condensates in *TRIM37Δ* cells. **b**, Immunoblot showing that PLK4 protein levels are not altered in *TRIM37Δ* cells. **c**, Immunofluorescence images of *TRIM37Δ* cells that lack centrioles owing to treatment with centrinone, showing centrosome components in an array of small condensates. **d**, Diagram highlighting the differences between the single large condensate found in *TRIM37Δ* cells with centrioles (top) and the small condensates in cells that lack centrioles owing to treatment with centrinone (bottom). **e**, Images of centrinone-treated *TRIM37Δ* cells, showing in situ CEP192 tagged with mNeonGreen (mNG) and the transgene-expressed, red fluorescent microtubule-binding domain of MAP4 protein (mRuby–MAP4–MBD). Times are minutes after NEBD. **f**, Control or *TRIM37Δ RPE1* cells with in situ tagged CEP192, treated with centrinone to inhibit PLK4 activity (top) or after inducible knock out of *PLK4* (bottom). Immunofluorescence (left) and plots of relative proliferation (right) show that foci formation and improved proliferation of centrinone-treated *TRIM37Δ* cells require PLK4 protein. Error bars show standard deviation (s.d.; $n = 3$). **g**, Analysis of mitotic duration for the conditions in **f**. Error bars show 95% confidence intervals. NS, not significant. **h**, Left, images showing localization to the centrosome of FLAG-tagged wild-type TRIM37 expressed in *TRIM37Δ* RPE1 cells. Middle, diagram showing the mutations investigated in the ligase and TRAF domains of TRIM37. Right, graph plotting the percentage of cells with condensates after expression in *TRIM37Δ* cells of FLAG-tagged wild-type or mutant TRIM37 proteins. **i**, Analysis of the interaction of TRIM37 variants with PLK4 following coexpression and PLK4 immunoprecipitation. The low expression of wild-type TRIM37 prompted us to use ligase-mutant TRIM37 in this analysis. IP, immunoprecipitation. **j**, The ubiquitination of PLK4 depends on the ligase activity of TRIM37 (observed following coexpression). α-Tubulin serves as a loading control for the input in **i**, **j**. Scale bars, 10 μm. For gel source data, see Supplementary Fig. 1. For details on statistics, see Methods; unpaired *t*-tests assuming equal standard deviation were performed.

smaller condensates that contained a larger subset of centrosomal components[13] (Fig. 2c, d and Extended Data Fig. 3d). These small condensates functioned as robust microtubule-organizing centres in both interphase and mitosis[13] (Fig. 2e and Extended Data Fig. 3f). Although it is not clear why condensates in interphase centrinone-treated *TRIM37Δ* cells recruit multiple centrosomal components, their ability to nucleate microtubules explains why reducing TRIM37 levels improves mitosis in centrinone (Fig. 1a, b, f and Extended Data Fig. 1d, e).

These data suggest that TRIM37 prevents PLK4 from self-assembling into condensates that can recruit other centrosomal components and function as ectopic microtubule-nucleating centres. This model predicts a requirement for catalytically inhibited PLK4 to form the arrays of smaller microtubule-generating condensates. To test this, we inducibly knocked out *PLK4* in control and *TRIM37Δ* cells (*iPLK4 KO*) (Extended Data Fig. 3g). Similar to centrinone treatment, induced *PLK4* knockout resulted in centrosome loss. However, in the absence of PLK4, *TRIM37Δ* did not cause formation of ectopic assemblies containing centrosomal proteins, nor improve acentrosomal division (Fig. 2f, g and Extended Data Fig. 3h). Thus, in *TRIM37Δ* cells, catalytically inhibited PLK4 acts as a scaffold for the formation of

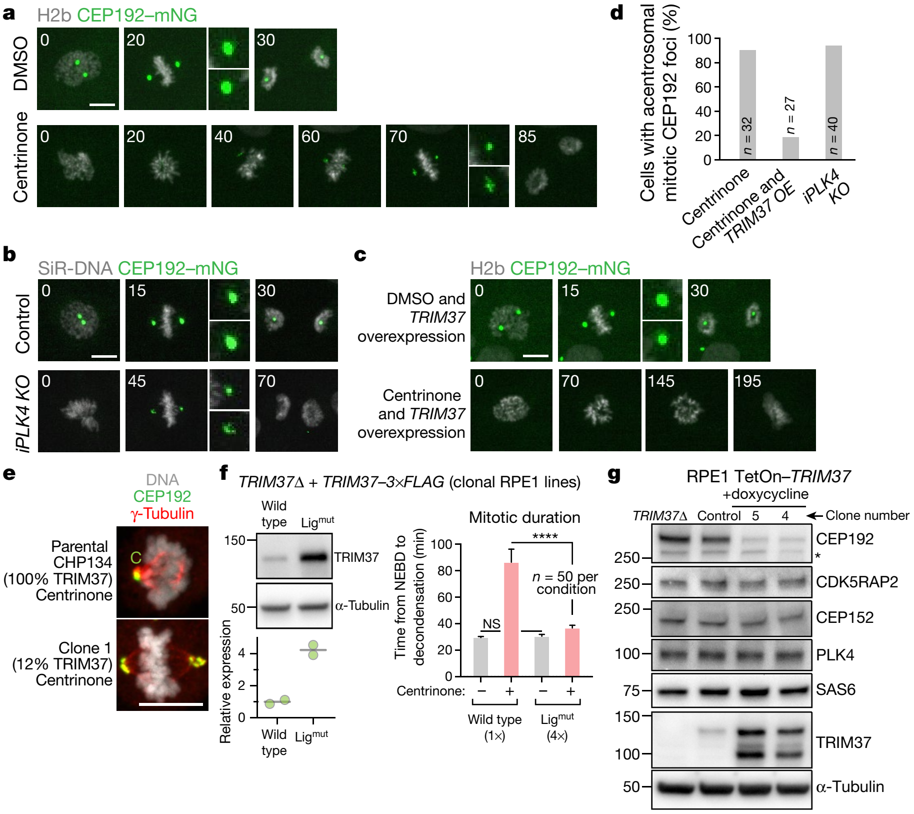


Fig. 3 | Acentrosomal spindle assembly and coalescence of pericentriolar material is suppressed by elevated TRIM37 levels in PLK4-inhibited cells. **a**, Images of mitosis in DMSO- or centrinone-treated RPE1 cells expressing in situ mNG-tagged CEP192. Acentrosomal spindle assembly in centrinone-treated cells is accompanied by the coalescence of CEP192 into foci at the spindle poles. Times are minutes after NEBD. **b**, Images of mitosis in control and *PLK4* knockout cells expressing in situ mNG-tagged CEP192, showing that the formation of foci at the spindle poles does not require PLK4 protein. Times are minutes after NEBD. **c**, Images of mitosis in TRIM37-overexpressing cells with in situ mNG-tagged CEP192 following treatment with DMSO or centrinone. Elevated TRIM37 expression suppresses the coalescence of pericentriolar material components and frequently results in cells exiting mitosis without segregating their chromosomes. Times are minutes after NEBD. **d**, Frequency of CEP192 coalescence into mitotic foci for the indicated conditions. **e**, Images of centrinone-treated mitotic CHP134 neuroblastoma parental cells or a clone with reduced TRIM37 expression (Fig. 1f). Acentrosomal foci are found only at spindle poles following TRIM37 reduction in this *TRIM37*-amplified (parental) cell line. C, centrosome. **f**, Ligase activity is required for cells with raised TRIM37 levels to exhibit increased sensitivity to PLK4 inhibition. Left, wild-type or ligase-inactive TRIM37 was expressed in *TRIM37Δ* RPE1 cells and clonal lines were isolated. Wild-type TRIM37 was expressed at a level comparable to that of endogenous TRIM37 (data not shown); ligase-inactive TRIM37 was expressed at a roughly fourfold higher level. Right, fourfold overexpression of ligase-inactive TRIM37 suppresses, rather than enhances, mitotic defects following treatment with centrinone. $^{****}P < 0.0001$. **g**, Effect of elevated TRIM37 expression on the indicated centrosomal components. CEP192 levels declined substantially whereas other tested components were not greatly affected. The asterisk marks a background band. α-Tubulin serves as a loading control in **f**, **g**. Scale bars, 10 μm. For gel source data, see Supplementary Fig. 1. For details on statistics, see Methods; unpaired *t*-tests assuming equal standard deviation were performed.

condensates that improve acentrosomal division by acting as ectopic microtubule-generating centres. Under low-ionic-strength conditions, purified PLK4 self-assembles into spherical condensates[17–19]. Our observations suggest that, in cells, this intrinsic property of PLK4 may be held in check by TRIM37.

TRIM37 ubiquitinates PLK4

TRIM37 is a tripartite motif ubiquitin ligase that has been localized to peroxisomes[20], but for which no centrosomal localization has been reported. We find that, in *TRIM37Δ* cells, epitope-tagged wild-type TRIM37 localizes to centrosomes and prevents the formation of PLK4 condensates (Fig. 2h and Extended Data Fig. 4a, c). TRIM37 has an RBCC (RING, B-box, coiled-coil) ubiquitin ligase domain and a TRAF domain that is predicted to mediate protein–protein interactions[21]. Point mutations engineered to disrupt ligase activity[22] or interactions between the TRAF domain and ligands[23,24] prevent the suppression of condensate formation in *TRIM37Δ* cells (Fig. 2h and Extended Data Fig. 4a–c). Expression of ligase-inactive TRIM37 was increased (Extended Data Fig. 4a), suggesting autoregulation by ligase activity. Notably, ligase-inactive TRIM37 localized to PLK4 condensates and centrosomes, in a TRAF-domain-dependent fashion (Extended Data Fig. 4c). As ligase-inactive mutants can act as substrate traps, the presence of ligase-inactive TRIM37 in condensates that contain PLK4 suggested that TRIM37 and PLK4 might interact. Supporting this idea, ligase-inactive TRIM37 associated with PLK4 when coexpressed in human cells, and this interaction was reduced in a TRAF-domain mutant (Fig. 2i and Extended Data Fig. 4d). Coexpression of tagged ubiquitin revealed TRIM37-dependent ubiquitination of PLK4 (Fig. 2j), without apparent reduction of PLK4 levels, consistent with regulation of PLK4 self-assembly and not stability (Fig. 2a, b). Although high TRIM37 expression in MCF7 cells has been reported to modulate gene expression by monoubiquitinating histone H2A[22], there was no reduction in monoubiquitinated H2A in *TRIM37Δ* RPE1 cells (Extended Data Fig. 4e), and transcript levels of *PLK4* and other centrosomal components were not altered by changes in TRIM37 expression (Extended Data Fig. 4f–h). Thus, TRIM37 acts to prevent the self-assembly of PLK4, rather than controlling its expression.

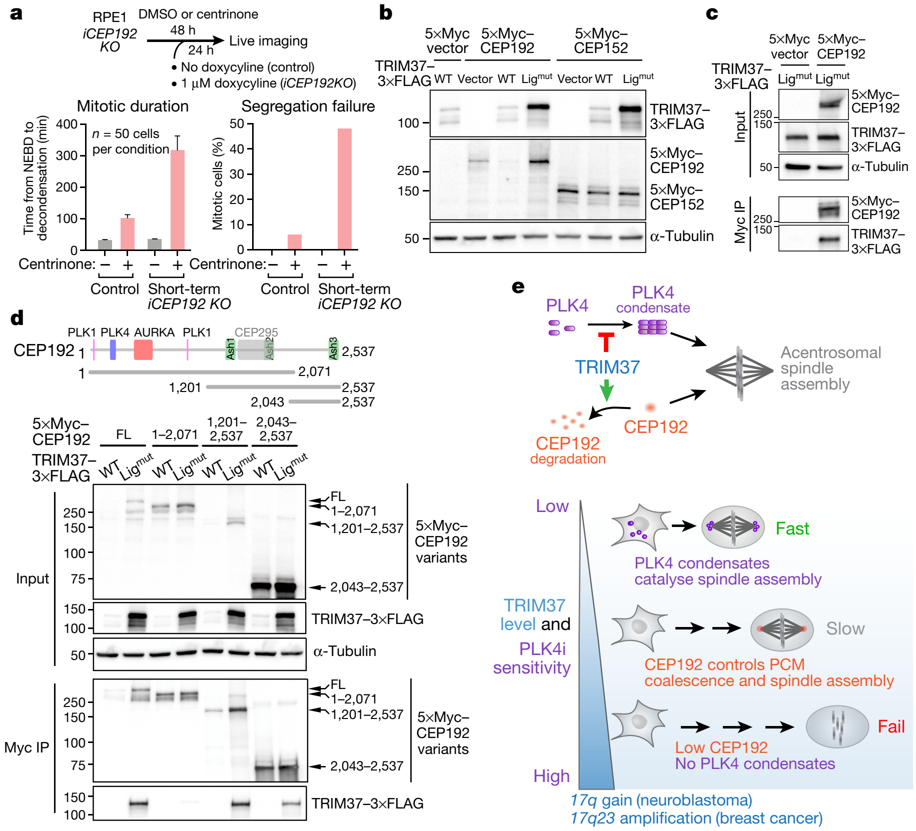


Fig. 4 | Elevated TRIM37 levels lead to a reduction in CEP192 that confers enhanced sensitivity to PLK4 inhibition. a, Top, approach for partially inhibiting CEP192 using a short-term inducible knockout (*iCEP192 KO*). Bottom, graphs plot mitotic duration and the percentage of cells with segregation failure. The short-term inducible *CEP192* knockout does not affect mitosis in DMSO-treated cells, but substantially enhances mitotic defects in centrinone-treated cells. **b**, Coexpression with wild-type (WT) or ligase-inactive TRIM37 shows that levels of CEP192, but not CEP152, protein are controlled by TRIM37's ligase activity. α-Tubulin serves as an input loading control. **c**, Interaction analysis showing that coexpressed ligase-inactive TRIM37 associates with CEP192. **d**, Top, diagram highlighting key interaction sites in CEP192. Bottom, the input blot shows the effects of coexpressed TRIM37 (wild-type or ligase-inactive) on the stability of CEP192 fragments; the immunoprecipitation blot assesses the association of CEP192 fragments with ligase-mutant TRIM37. When CEP192 cannot interact with TRIM37 because its C-terminus is deleted (CEP192 1–2,071), levels of the CEP192 protein are not affected by coexpression of wild-type TRIM37. FL, full length. **e**, Model depicting how TRIM37 exerts bidirectional control over acentrosomal mitosis following inhibition of PLK4. Top, two ligase-activity-dependent functions of TRIM37 are to prevent PLK4 from self-assembling into condensates that nucleate microtubules, and to target CEP192 for degradation. Bottom, when TRIM37 levels are low, PLK4 forms condensates that catalyse robust acentrosomal spindle assembly. When TRIM37 levels are normal, TRIM37 prevents PLK4 from forming condensates; after mitotic entry, foci containing pericentriolar material (PCM) components coalesce coincident with slow acentrosomal spindle assembly. When TRIM37 levels are high, CEP192 levels are reduced and there are no PLK4 condensates; consequently, acentrosomal spindle assembly fails. Amplification of the genomic region containing *TRIM37* in neuroblastoma and a subset of breast cancers highlights the potential for synthetic lethality with inhibition of PLK4 in specific cancer contexts. For gel source data, see Supplementary Fig. 1.

TRIM37 and acentrosomal mitotic foci

We next focused on understanding why elevated TRIM37 levels lead to mitotic failure in PLK4-inhibited cells (Fig. 1a, b, f and Extended Data Fig. 1d, e). As inducible *PLK4* knockout and centrinone treatment produce mitotic defects of similar magnitude (Fig. 2f, g), sensitivity to PLK4 inhibition caused by elevated TRIM37 expression cannot be due to TRIM37 limiting the formation of PLK4-scaffolded ectopic microtubule-organizing centres; if it were, then *PLK4* knockout should produce more severe mitotic defects than centrinone treatment. To address how elevated TRIM37 expression enhances sensitivity to PLK4 inhibition, we monitored cells with an in situ tagged fluorescent fusion of the pericentriolar material protein CEP192. Although we detected no concentration of centrosomal proteins during interphase in centrinone-treated cells[4], following nuclear envelope breakdown (NEBD), CEP192 and a collection of centrosomal proteins did gradually coalesce to form foci positioned at the spindle poles (Fig. 3a and Extended Data Fig. 5a–c). Formation of these foci was observed with the same timing and frequency after induced *PLK4* knockout (Fig. 3b, d and Extended Data Fig. 5d, f). In TRIM37-overexpressing RPE1 cells treated with centrinone, foci containing CEP192 and other centrosomal components failed to form (Fig. 3c, d and Extended Data Fig. 5e, f). In centrinone-treated CHP134 neuroblastoma cells with *TRIM37* amplification, no acentrosomal mitotic foci were observed unless TRIM37 levels were reduced (Fig. 3e). Notably, the presence of these acentrosomal mitotic foci correlated with bipolar spindle formation and chromosome segregation (Fig. 1a, b and Extended Data Fig. 5f). Thus, elevated TRIM37 levels inhibit the formation of mitotic pericentriolar material foci that occurs coincident with acentrosomal spindle assembly.

TRIM37 ligase regulates CEP192 stability

To define the molecular mechanism by which increased TRIM37 levels enhance sensitivity to PLK4 inhibition, we first assessed whether TRIM37 ligase activity is important. A roughly fourfold overexpression of ligase-dead TRIM37 failed to enhance mitotic defects following centrinone treatment (Fig. 3f), unlike a comparable overexpression of wild-type TRIM37 (Fig. 1a, b). In fact, ligase-dead TRIM37 improved mitotic outcomes following centrinone treatment because it phenocopied loss of TRIM37, in terms of promoting the formation of PLK4 condensates that serve as ectopic microtubule-generating centres (Fig. 2h and Extended Data Fig. 4c). Thus, ligase activity is required for elevated TRIM37 levels to render mitosis sensitive to PLK4 inhibition. Immunoblotting of a panel of centrosome components in control, *TRIM37Δ* and TRIM37-overexpressing cells revealed that CEP192 levels were greatly decreased when TRIM37 was overexpressed, whereas levels of other analysed components were unchanged (Fig. 3g). Notably, TRIM37 overexpression did not substantially affect the transcriptome (Extended Data Fig. 5g). The effect of TRIM37 on CEP192 protein levels was enhanced following centrinone treatment (Extended Data Fig. 5h), suggesting that centrosomes protect CEP192 from TRIM37-dependent degradation. This protection could be direct (mediated by localization of CEP192 to centrosomes) or indirect (resulting from prolonged mitosis in the absence of centrosomes). Centrosome-dependent protection of CEP192 probably explains why mitosis in cells with centrosomes is not affected by increased TRIM37 levels (Fig. 1a, b and Extended Data Fig. 1d, e).

Given that elevated TRIM37 levels reduce CEP192 protein levels and selectively disrupt acentrosomal mitosis, we next tested whether reducing CEP192 levels by a different means also disrupts acentrosomal but not centrosomal mitosis. As CEP192 is essential, it is not possible to delete the gene encoding it to assess centrinone sensitivity. Instead, we partially inhibited CEP192 using a short-term conditional knockout that was well-tolerated in mock-treated cells (Fig. 4a and Extended Data Fig. 6a). Partial CEP192 inhibition selectively disrupted mitosis in centrinone-treated cells (Fig. 4a), analogous to TRIM37 overexpression (Fig. 1a, b and Extended Data Fig. 1d, e). A similar result was observed using a short hairpin RNA (shRNA) in a CHP134 clonal line with reduced TRIM37 expression (Extended Data Fig. 6b). These functional data indicate that elevated TRIM37 activity confers enhanced sensitivity to PLK4 inhibition by reducing CEP192 levels. Consistent with this model, coexpression of TRIM37 in HEK293 cells reduced CEP192 but not CEP152 levels in a ligase-activity-dependent manner (Fig. 4b), and CEP192 coimmunoprecipitated with ligase-inactive TRIM37 (Fig. 4c).

CEP192 is a multifunctional scaffold that binds PLK4 to control centriole duplication[25–27], and to the mitotic kinases PLK1 and Aurora A to control the assembly of pericentriolar material[28]. Coexpression of CEP192 fragments with TRIM37 indicated that a 495-amino-acid carboxy-terminal region of CEP192—distinct from previously characterized CEP192-interaction regions—bound robustly to TRIM37, was ubiquitinated in a TRIM37-dependent manner, and was required for TRIM37-dependent degradation (Fig. 4d and Extended Data Fig. 6c). These data indicate that CEP192, as with PLK4, is a direct TRIM37 target. However, in contrast with PLK4, CEP192 protein levels are controlled by TRIM37 ligase activity, especially when centrosomes are absent. Thus, raised TRIM37 levels confer sensitivity to PLK4 inhibition by causing a reduction in CEP192 levels.

Xenograft sensitivity to PLK4 inhibition

Although centrinone is highly selective towards PLK4 and effective in culture, its pharmacokinetic profile has precluded its use in tumour models (data not shown). We therefore used inducible shRNA to test the sensitivity of CHP134 xenografts to PLK4 inhibition. Induction of a *PLK4* shRNA mimicked centrinone treatment, causing a reduction in centrosome number and rapid loss of CHP134 viability (Extended Data Fig. 7a–d). Xenograft tumours were generated using two CHP134 *PLK4* shRNA clones in nude mice, with the feed switched to induce shRNA expression (Extended Data Fig. 7e). *PLK4* shRNA induction suppressed tumour growth for both clones (Extended Data Fig. 7f). To assess whether the magnitude of the reduction of CHP134 xenograft tumour growth was influenced by the amount of TRIM37 present, we analysed tumour formation by CHP134 parental cells and two derived clonal lines with reduced expression of TRIM37 (Fig. 1f, clones 1, 2). CHP134 clones expressing low TRIM37 levels exhibited poorer tumour growth than parental CHP134 cells (data not shown); this result was reminiscent of prior work on TRIM37 as an oncoprotein in breast cancer[22]. As tumour growth was influenced by TRIM37 levels, it was not feasible to analyse the effect of different levels of TRIM37 expression. Nevertheless, our xenograft tumour experiments highlight the potential of PLK4 inhibition as a therapeutic strategy for neuroblastoma and potentially also other cancers with high *TRIM37* expression.

Conclusion

We have shown that the centrosomal ubiquitin ligase TRIM37 functions as a rheostat that controls cell division in the presence of chemical inhibition of PLK4 (Fig. 4e). The molecular mechanisms by which low versus high TRIM37 expression influences mitosis in PLK4-inhibited cells are surprisingly distinct (Fig. 4e). Neuroblastoma and breast cancer cells with genomic amplification of *TRIM37* are highly sensitive to PLK4 inhibition; an independent effort reached a similar conclusion for 17q23-amplified breast cancers[29]. For these, as well as for other cancer types with amplification of the *TRIM37* locus, inhibition of PLK4 offers a new approach for selectively triggering mitotic failure. The case for PLK4 inhibition as a therapeutic strategy is particularly compelling for neuroblastoma, which is the most common extracranial solid paediatric cancer and accounts for around 13% of paediatric deaths from cancer[30,31]. About half of neuroblastoma is high risk[31,32], and nearly 80% of the high-risk cases have a gain of 17q (ref. [6]). At present the mortality from high-risk neuroblastoma is roughly 50%, and survivors suffer treatment-related morbidity[33]. Our results highlight the importance of developing new highly selective PLK4 inhibitors with improved properties for testing in preclinical and clinical studies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2710-1.

1. Petry, S. Mechanisms of mitotic spindle assembly. *Annu. Rev. Biochem*. **85**, 659–683 (2016).
2. Nigg, E. A. & Holland, A. J. Once and only once: mechanisms of centriole duplication and their deregulation in disease. *Nat. Rev. Mol. Cell Biol*. **19**, 297–312 (2018).
3. Zitouni, S., Nabais, C., Jana, S. C., Guerrero, A. & Bettencourt-Dias, M. Polo-like kinases: structural variations lead to multiple functions. *Nat. Rev. Mol. Cell Biol*. **15**, 433–452 (2014).
4. Wong, Y. L. et al. Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science* **348**, 1155–1160 (2015).
5. Bulavin, D. V. et al. Amplification of PPM1D in human tumors abrogates p53 tumor-suppressor activity. *Nat. Genet*. **31**, 210–215 (2002).
6. Ho, N. et al. Delineation of the frequency and boundary of chromosomal copy number variations in paediatric neuroblastoma. *Cell Cycle* **17**, 749–758 (2018).
7. Li, J. et al. Oncogenic properties of PPM1D located within a breast cancer amplification epicenter at 17q23. *Nat. Genet*. **31**, 133–134 (2002).
8. Liu, Y. et al. Targeting 17q23 amplicon to overcome the resistance to anti-HER2 therapy in HER2+ breast cancer. *Nat. Commun*. **9**, 4718 (2018).
9. Woodruff, J. B., Wueseke, O. & Hyman, A. A. Pericentriolar material structure and dynamics. *Phil. Trans. R. Soc. Lond. B* **369**, 20130459 (2014).
10. Oegema, K., Davis, R. L., Lara-Gonzalez, P., Desai, A. & Shiau, A. K. CFI-400945 is not a selective cellular PLK4 inhibitor. *Proc. Natl Acad. Sci. USA* **115**, E10808–E10809 (2018).
11. Bazzi, H. & Anderson, K. V. Acentriolar mitosis activates a p53-dependent apoptosis pathway in the mouse embryo. *Proc. Natl Acad. Sci. USA* **111**, E1491–E1500 (2014).
12. Khodjakov, A. & Rieder, C. L. Centrosomes enhance the fidelity of cytokinesis in vertebrates and are required for cell cycle progression. *J. Cell Biol*. **153**, 237–242 (2001).
13. Meitinger, F. et al. 53BP1 and USP28 mediate p53 activation and G1 arrest after centrosome loss or extended mitotic duration. *J. Cell Biol*. **214**, 155–166 (2016).
14. Sir, J. H. et al. Loss of centrioles causes chromosomal instability in vertebrate somatic cells. *J. Cell Biol*. **203**, 747–756 (2013).
15. Ma, X. et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371–376 (2018).

16. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
17. Montenegro Gouveia, S. et al. PLK4 is a microtubule-associated protein that self-assembles promoting *de novo* MTOC formation. *J. Cell Sci*. **132**, jcs219501 (2019).
18. Park, J. E. et al. Phase separation of Polo-like kinase 4 by autoactivation and clustering drives centriole biogenesis. *Nat. Commun*. **10**, 4959 (2019).
19. Yamamoto, S. & Kitagawa, D. Self-organization of Plk4 regulates symmetry breaking in centriole duplication. *Nat. Commun*. **10**, 1810 (2019).
20. Brigant, B., Metzinger-Le Meuth, V., Rochette, J. & Metzinger, L. TRIMming down to TRIM37: relevance to inflammation, cardiovascular disorders, and cancer in MULIBREY nanism. *Int. J. Mol. Sci*. **20**, 67 (2018).
21. Park, H. H. Structure of TRAF family: current understanding of receptor recognition. *Front. Immunol*. **9**, 1999 (2018).
22. Bhatnagar, S. et al. TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. *Nature* **516**, 116–120 (2014).
23. Hu, M. et al. Structural basis of competitive recognition of p53 and MDM2 by HAUSP/USP7: implications for the regulation of the p53-MDM2 pathway. *PLoS Biol*. **4**, e27 (2006).
24. Sheng, Y. et al. Molecular recognition of p53 and MDM2 by USP7/HAUSP. *Nat. Struct. Mol. Biol*. **13**, 285–291 (2006).
25. Kim, T. S. et al. Hierarchical recruitment of Plk4 and regulation of centriole biogenesis by two centrosomal scaffolds, Cep192 and Cep152. *Proc. Natl Acad. Sci. USA* **110**, E4849–E4857 (2013).
26. Park, S. Y. et al. Molecular basis for unidirectional scaffold switching of human Plk4 in centriole biogenesis. *Nat. Struct. Mol. Biol*. **21**, 696–703 (2014).
27. Sonnen, K. F., Gabryjonczyk, A. M., Anselm, E., Stierhof, Y. D. & Nigg, E. A. Human Cep192 and Cep152 cooperate in Plk4 recruitment and centriole duplication. *J. Cell Sci*. **126**, 3223–3233 (2013).
28. Joukov, V. & De Nicolo, A. Aurora-PLK1 cascades as key signaling modules in the regulation of mitosis. *Sci. Signal*. **11**, eaar4195 (2018).
29. Yeow, Z. Y. et al. Targeting TRIM37-driven centrosome dysfunction in 17q23-amplified breast cancer. *Nature* https://doi.org/10.1038/s41586-020-2690-1 (2020).
30. Cheung, N. K. & Dyer, M. A. Neuroblastoma: developmental biology, cancer genomics and immunotherapy. *Nat. Rev. Cancer* **13**, 397–411 (2013).
31. Louis, C. U. & Shohet, J. M. Neuroblastoma: molecular pathogenesis and therapy. *Annu. Rev. Med*. **66**, 49–63 (2015).
32. Maris, J. M. Recent advances in neuroblastoma. *N. Engl. J. Med*. **362**, 2202–2211 (2010).
33. Oeffinger, K. C. et al. Chronic health conditions in adult survivors of childhood cancer. *N. Engl. J. Med*. **355**, 1572–1582 (2006).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Antibodies

Antibodies against CEP192 (amino acids 1–211; used at 0.5 μg ml^{-1} for immunofluorescence and immunoblotting), SAS6 (amino acids 501–657; used at 0.5 μg ml^{-1} for immunofluorescence) and PLK4 (amino acids 814–970; used at 1 μg ml^{-1} for immunofluorescence) have previously been described[4]. The following antibodies were purchased from commercial sources, with their working concentrations indicated in parentheses: anti-TRIM37 (1:2,000 for immunoblotting; catalogue number A301-174A, Bethyl Laboratories); anti-PLK4 (1:500 for immunoblotting; clone 6H5, MABC544, Merck Millipore); anti-CEP152 (1:1,000 for immunofluorescence; ab183911, Abcam); anti-CEP152 (1:1,000 for immunoblotting; A302-479A-T, Bethyl Laboratories); anti-CDK5RAP2 (1:1,000 for immunofluorescence; ab86340, Abcam); anti-CDK5RAP2; (1:1,000 for immunoblotting; A300-554A-T, Bethyl Laboratories); anti-γ-tubulin (1:1,000 for immunofluorescence; GTU-88, Sigma-Aldrich); anti-pericentrin (1 μg ml^{-1} for immunofluorescence; ab4448, Abcam), anti-CPAP (1:400 for immunofluorescence; 11517-1-AP, Proteintech); anti-CCDC14 (1:100 for immunofluorescence; PA5-31759, Thermo Fisher Scientific); anti-CEP63 (1:100 for immunofluorescence; 06-1292, Merck-Millipore); anti-KIAA0753 (1:500 for immunofluorescence; HPA023494, Sigma-Aldrich); anti-PCM-1 (1:400 for immunofluorescence; 5259, Cell Signaling Technology); anti-CEP135 (1:500 for immunofluorescence; ab75005, Abcam); anti-α-tubulin (1:5000 for immunoblotting; DM1A, Sigma-Aldrich); anti-FLAG (1:1,000 for immunoblotting; F1804, Sigma-Aldrich); anti-Myc (1:5,000 for immunoblotting; monoclonal antibody 9E10, M4439, Sigma-Aldrich); and anti-HA (1:500 for immunoblotting; monoclonal antibody 16B12, BioLegend). Secondary antibodies were purchased from Jackson ImmunoResearch and GE Healthcare.

Cell lines

All cell lines used here are described in Extended Data Table 1. RPE1 (hTERT RPE-1), CHP212, IMR32, SK-N-F1, BT474, BT549, MCF7, MDA-MB-231 and HepG2 cell lines were obtained from the American Type Culture Collection (ATCC). The CHP134 cell line was obtained from Sigma-Aldrich (ECACC general collection). The KPNYN line was a gift from P. Zage. Cell lines obtained from the ATCC and Sigma were cultured as recommended. Each growth medium was supplemented with 100 IU ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. All cell lines except FreeStyle 293-F cells were maintained at 37 °C and 5% CO$_2$. FreeStyle 293-F cells were maintained at 37 °C and 8% CO$_2$ in air on an orbital shaker platform rotating at 125 r.p.m. All cell lines have been tested for mycoplasma contamination. To inhibit PLK4 and deplete centrosomes, cells were treated with centrinone for the indicated times[4]. Centrinone was diluted from a 1 mM stock; all treatments were at a final concentration of 150 nM centrinone and 0.015% DMSO; control treatments were 0.15% DMSO

The RPE1 *TRIM37Δ* cell line has previously been described[13]. *TRIM37Δ* and *USP28Δ* knockouts in the RPE1 CEP192–mNeonGreen background were generated as described[13]. In brief, double-stranded oligonucleotides for specific guide RNAs targeting *USP28* (TGAGCGTTTAGTTTCTGCAG) or *TRIM37* (CTCCCCAAAGTGCACACTGA) were cloned into PX459 (a gift from F. Zhang; Addgene plasmid 48139; http://n2t.net/addgene:48139; Resource Identification Portal (RRID; https://scicrunch.org/resources) identification code Addgene_48139)[34]. RPE1 cells were plated in 10-cm plates at 500,000 cells per plate the day before transfection. Cells were transfected with plasmid using Lipofectamine 3000 according to the manufacturer's instructions (ThermoFisher). Two days after transfection, 100 nM centrinone was added. After 10 days, centrinone-resistant RPE1 cells were plated at a density that supported direct picking of clones from 10-cm plates. One week after re-plating, multiple colonies were observed at a density that supported direct picking of clones. Gene knockout was determined by genotyping of the sequence surrounding the CRISPR cut site[35] and/or by immunoblotting.

CHP134 cell lines with variable amounts of TRIM37 expression were generated using CRISPR–Cas9. CHP134 cells were plated in six-well dishes at 200,000 cells per well the day before transfection. On the day of transfection, a synthetic CRISPR RNA (crRNA) targeting exon 5 of *TRIM37* (CTCCCCAAAGTGCACACTGA) was hybridized with synthetic transactivating crRNA (tracrRNA), assembled with Cas9 protein into ribonucleoproteins (RNPs), and transfected into CHP134 cells using Lipofectamine RNAiMAX according to the manufacturer's instructions (Thermo Fisher Scientific). To obtain clonal lines, single cells were plated into 96-well plates and expanded. Targeting of *TRIM37* was determined by genotyping of the region surrounding the CRISPR cut site[35] and immunoblotting.

RPE1 CEP192–mNeonGreen cells were generated using CRISPR–Cas9 in combination with recombinant adeno-associated virus (rAAV)-mediated delivery of the repair construct as described[36]. The guide RNA (gRNA) was designed to cut close to the stop codon of *CEP192* (CGACTAATTGGTGAAGCTCT) and cloned into PX459 (ref. [34]). The repair construct was cloned into the pSEPT plasmid and contains the left and right flanking regions of the gRNA target site (respectively 960 and 672 base pairs); the monomeric NeonGreen (mNeonGreen) coding sequence, for C-terminal fusion to CEP192, and the neomycin-resistance gene aminoglycoside phosphotransferase from transposon Tn5 was cloned between the left and right homology arms. The expression of the neomycin-resistance gene is linked to endogenous CEP192–NeonGreen expression through a P2A sequence.

The following transgenes were stably integrated into the genome using lentiviral constructs (see Extended Data Table 2): histone H2B fused to monomeric red fluorescent protein (*H2B–mRFP*; *EF1alpha* promoter); the microtubule-associated protein 4 (MAP4) microtubule-binding domain (MBD) fused to monomeric Ruby2 (*mRuby2–MAP4–MBD*; *EF1alpha* promoter; neomycin-resistance gene); TRIM37 with the C18R mutation fused to mNeonGreen (*TRIM37-C18R–mNeonGreen*; *hPGK* promoter; blasticidin-resistance gene) and TRIM37 with three FLAG epitope sequences (*TRIM37–3×FLAG*: wild-type, *C18R*, *W373A* and *C18R W373A*; *UbC* promoter; neomycin-resistance gene). Cell lines with inducible overexpression of TRIM37 were generated by sequential lentiviral integration of *TetOn3G* (inducible transactivator protein; *hPGK* promoter) and *TRIM37* (doxycycline-inducible *TRE3GS* promoter). Cell lines for inducible knockout of *PLK4* or *CEP192* were generated by sequential lentiviral integration of Cas9 (Edit-R inducible lentiviral Cas9; Dharmacon) and a *PLK4* or *CEP192* gRNA-expressing plasmid, based on the lentiGuide-Puro plasmid[37]. The *PLK4* gRNA (TCATATTACGAGTCAGTAGG) targets exon 5 in the kinase-domain-coding region. The *CEP192* gRNA (AGGGAGTGTCCGAGTGCCCG) targets exon 19. The lentiGuide-Puro was a gift from F. Zhang (Addgene plasmid 52963; http://n2t.net/addgene:52963; RRID Addgene_52963). *TRIM37* and Cas9 expression were induced with 1 μg ml^{-1} doxycycline. Viral particles were generated by transfecting the lentiviral construct into HEK-293T cells using Lenti-X Packaging Single Shots (Takara Bio USA). Forty-eight hours after transfection, virus-containing culture supernatant was collected and added to the growth medium of cells in combination with 2.5–8 μg ml^{-1} polybrene (EMD Millipore). Populations of each cell line were selected by fluorescence-activated cell sorting (FACS) or antibiotics (blasticidin, 5 μg ml^{-1}; neomycin, 400 μg ml^{-1}; puromycin, 10 μg ml^{-1} for RPE1 cells). Single clones were isolated in 96-well plates. The lentiviral vector expressing *CEP192* shRNA (GAGGCATCAGTTAATACTGAT) was purchased from Dharmacon. The *PLK4* shRNA (CAGTATAAGTGGTAGTTTA) was expressed from an integrated lentiviral vector and expressed from a doxycycline-inducible promoter (construct name V3SH11252-225330936 piSMART hEF1a/TurboGFP inducible). Single clones were selected for tumour xenograft experiments.

Immunofluorescence analysis

For immunofluorescence, 10,000 cells per well were seeded into 96-well plates one day before fixation. Cells were fixed in 100 µl ice-cold methanol for 7 min at −20 °C. Cells were washed twice with washing buffer (phosphate-buffered saline (PBS) containing 0.1% Triton X-100) and blocked with blocking buffer (PBS containing 2% bovine serum albumin (BSA), 0.1% Triton X-100 and 0.1% NaN_3) overnight. After blocking, cells were incubated for 1–2 h with primary antibody in fresh blocking buffer (concentrations as indicated above). Cells were washed three times with washing buffer, before a 1-h incubation with the secondary antibody and DNA staining with Hoechst 33342 dye. Finally, cells were washed three times with washing buffer before inspection. Images were acquired on a CV7000 spinning disk confocal system (Yokogawa Electric) equipped with a ×40 (numerical aperture (NA) 0.95) or a ×60 (water, NA 1.2) U-PlanApo objective and a 2,560 × 2,160 pixel sCMOS camera (Andor). Image acquisition was performed using CV7000 software.

Live-cell imaging

Live-cell imaging was performed with the CQ1 spinning disk confocal system (Yokogawa Electric) equipped with a ×40 NA 0.95 U-PlanApo objective and a 2,560 × 2,160 pixel sCMOS camera (Andor) at 37 °C and 5% CO_2. Image acquisition and data analysis were performed using CQ1 software and ImageJ, respectively.

Cells were seeded into 96-well polystyrene plates at 6,000–10,000 cells per well, 24 h before imaging, unless indicated otherwise. Imaging conditions varied according to the experimental setup. For imaging of H2B–RFP, or of DNA with the DNA marker SiR–DNA, a 5 × 2 µm *z*-stack in the RFP or FarRed channel (25% power, 150 ms) was acquired of each field at 4- or 5-min intervals for 6–24 h. SiR–DNA was added 2 h before imaging at a working concentration of 0.5 µM. For imaging of CEP192–NeonGreen and/or mRuby–MAP4–MBP, an 8 × 1.2 µm *z*-stack in the green fluorescent protein (GFP) and/or RFP channel (50% power, 150 ms) was acquired of each field at 4- to 15-min intervals for 6–12 h. DMSO or centrinone treatment was conducted for three cell cycles before the start of imaging, unless noted otherwise; the duration of the cell cycle for RPE1 and neuroblastoma cell lines was measured by live imaging of each cell line and quantifying time from NEBD of a mother cell to NEBD of its daughters.

Proliferation and viability assays

For the passaging assays, cells were seeded into six-well plates in triplicate at 25,000 cells per well and treated with 150 nM centrinone or DMSO. At 96-h or 192-h intervals, cells were collected, counted and re-plated at 25,000 cells per well. Cell counting was performed using a TC20 automated cell counter (Bio-Rad).

For ATPlite viability assays, 550–750 cells were seeded into 96-well culture plates (Corning 3603) in 180 µl and incubated overnight at 37 °C and 5% CO_2. The following day, 20 µl medium containing 1.25 µM centrinone (or equal volume with DMSO as control) was added to each well to obtain a final concentration of 125 nM. After five days of incubation at 37 °C and 5% CO_2, 100 µl ATPlite (Perkin Elmer) was added before luminescence measurement with a Tecan Infinite M1000 Pro Multilabel microplate reader.

Immunoblotting

For immunoblotting, cells were cultured in 10-cm plates, collected at 50–80% confluence and lysed by sonication in RIPA buffer (Cell Signaling Technology) plus protease and phosphatase-inhibitor cocktail (Thermo Fisher Scientific). Cell extracts were stored at −80 °C until use. Before use, extract concentrations were normalized on the basis of a protein assay (Bio-Rad). For every sample, 20–30 µg of protein per lane were run on Mini-Protean gels (Bio-Rad), and transferred to polyvinylidene fluoride (PVDF) membranes using a TransBlot Turbo system (Bio-Rad). Blocking and antibody incubations were performed in Tris-buffered saline plus Tween-20 (TBS-T) with 5% non-fat dry milk. Detection was performed using horseradish peroxidase (HRP)-conjugated secondary antibodies (GE Healthcare) with WesternBright Sirius (Advansta) or SuperSignal West Femto (Thermo Fisher Scientific) substrates. Membranes were imaged on a ChemiDoc MP system (Bio-Rad).

Protein expression and immunoprecipitation

For coexpression and immunoprecipitation assays, FLAG-tagged TRIM37 and Myc-tagged PLK4, CEP192 or CEP152 (Extended Data Table 2) were expressed in different combinations in FreeStyle 293-F cells (Thermo Fisher Scientific). Complementary DNA constructs for *CEP192* and *CEP152* transient expression[25] were gifts from K. S. Lee. The empty 5×Myc plasmid, which was used as negative control, is a derivative of CS2P (Addgene 17095) and is designed for expression of C-terminally Myc-tagged proteins. Cell transfection was performed using FreeStyle MAX Reagent and OptiPRO SFM according to the manufacturer's guidelines (Thermo Fisher Scientific). Next, 20 ml of cells at 1×10^6 cells per ml were transfected with a total of 25 µg DNA constructs. Forty-three to forty-eight hours after transfection, cells were collected and washed with PBS. The cells were resuspended in lysis buffer (20 mM Tris/HCl pH 7.5, 150 mM NaCl, 1% Triton X-100, 5 mM EGTA, 1 mM dithiothreitol, 2 mM $MgCl_2$ and EDTA-free protease-inhibitor cocktail (Roche)) and lysed in an ice-cold sonicating water bath for 5 min. After 15-min centrifugation at 15,000*g* and 4 °C, whole-cell lysates were incubated with Pierce anti-Myc magnetic beads (Thermo Fisher Scientific) for 2 h at 4 °C. The beads were washed five times with lysis buffer and resuspended in SDS sample buffer. For immunoblotting, equal volumes of samples were run on Mini-Protean gels (Bio-Rad) and transferred to PVDF membranes using a TransBlot Turbo system (Bio-Rad). Blocking and antibody incubations were performed in TBS-T plus 5% non-fat dry milk or in TBS-T plus 5% BSA. Immunoblotting was performed as above.

Detection of protein ubiquitination

To detect ubiquitination of PLK4 and CEP192 by TRIM37, we expressed DNA constructs encoding Myc-tagged PLK4 or CEP192 along with haemagglutinin (HA)-tagged ubiquitin and FLAG-tagged TRIM37 in FreeStyle 293-F cells for 48 h. HA–ubiquitin was a gift from E. Yeh (Addgene plasmid 18712; http://n2t.net/addgene:18712; RRID Addgene_18712)[38]. The cells were lysed in 20 mM Tris/HCl pH 7.5, 150 mM NaCl, 1% Triton X-100, 5 mM EGTA, 1 mM dithiothreitol, 2 mM $MgCl_2$, EDTA-free protease-inhibitor cocktail (Roche) and 5 mM *N*-ethylmaleimide. Immunoprecipitation and immunoblotting were performed as above.

Statistical analysis

P-values were obtained from *t*-tests conducted using Prism v8 (GraphPad). For Figs. 2g, 3f and Extended Data Figs. 1d, 1g, 2g, 4e, 6b, unpaired *t*-tests assuming equal standard deviation were performed. For Extended Data Fig. 7f, unpaired *t*-tests with Welch's correction, which does not assume equal standard deviations, were performed at each time point; only the significantly different ($P < 0.05$) time points are marked. *P*-values are labelled as follows: NS, $P > 0.05$; $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

RNA-sequencing analysis

RNA was purified from three independent samples for each analysed cell line using an RNeasy Plus Mini Kit. RNA library synthesis and sequencing were performed by the Genomics Center of the Institute for Genomic Medicine at UC San Diego. Samples were sequenced on a HiSeq4000 platform (SR75). The sequencing data were aligned to the human genome UCSC hg19 annotation with STAR aligner[39]. Differential expression of genes was determined using DESeq2 (ref. [40]). Genes with low counts (an average of fewer than ten reads for triplicates) or outliers (highly skewed value in one of the triplicates) were excluded from the analysis. Results were visualized using Prism v8 and InteractiVenn[41].

CHP134 xenograft tumour analysis

Tumour xenografts were initiated by inoculation of CHP134 human neuroblastoma cells stably transduced with an inducible *PLK4* shRNA. Two independent clones (ODCL108 and ODCL109) were used to generate xenograft tumours in six-week old BALB/c nude female mice. Cells (1×10^7 per mouse) were suspended in 1:1 PBS:matrigel (Corning); 100 μl of the cell suspension was injected into the subcutaneous right flank. For the inducible *PLK4* shRNA, ODCL108 mice were randomized (two groups with nine mice each) and switched to irradiated control or doxycycline-containing (625 mg kg^{-1}) feed when tumours reached an average size of about 150 mm^3; ODCL109 mice exhibited slower tumour growth and were randomized (two groups with nine mice each) and switched when tumours reached an average size of about 100 mm^3. Tumour size was calculated by standard caliper measurement, using volume = ($width^2 \times length$)/2. All procedures related to mouse handling, care and treatment followed guidelines approved by the Institutional Animal Care and Use Committee (IACUC) of BioDuro, San Diego, following the guidance of the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). BioDuro's limit on conventional mouse xenograft tumour size is 2,000 mm^3. One control mouse from the P7 clone exceeded this limit on the final study day before all P7 mice were killed.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The RNA-sequencing data in Extended Data Figs. 4f, g, 5g have been deposited in the National Center for Biotechnology Information (NCBI)'s Gene Expression Omnibus (GEO)[42] and can be accessed with GEO accession number GSE148263 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE148263). Other data or materials are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

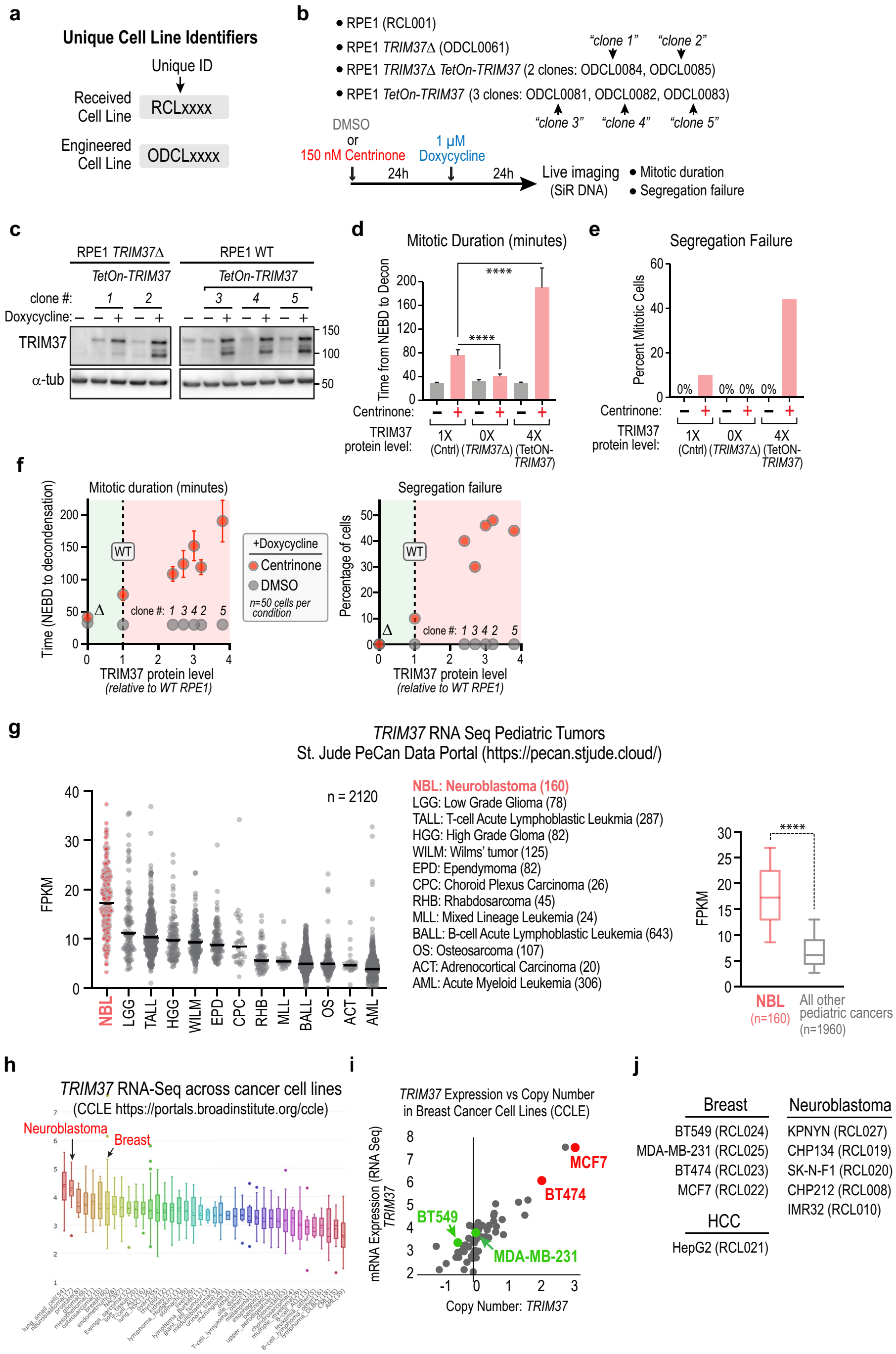
34. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protocols* **8**, 2281–2308 (2013).
35. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*. **42**, e168 (2014).
36. Kaulich, M. & Dowdy, S. F. Combining CRISPR/Cas9 and rAAV templates for efficient gene editing. *Nucleic Acid Ther*. **25**, 287–296 (2015).
37. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
38. Kamitani, T., Kito, K., Nguyen, H. P. & Yeh, E. T. Characterization of NEDD8, a developmentally down-regulated ubiquitin-like protein. *J. Biol. Chem*. **272**, 28557–28562 (1997).
39. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
40. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol*. **11**, R106 (2010).
41. Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169 (2015).
42. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. **30**, 207–210 (2002).
43. Bauer, M., Cubizolles, F., Schmidt, A. & Nigg, E. A. Quantitative analysis of human centrosome architecture by targeted proteomics and fluorescence imaging. *EMBO J*. **35**, 2152–2166 (2016).

Acknowledgements We dedicate this manuscript to the memory of T. Gahman, our colleague and friend, with whom we developed centrinone. This work was supported by a National Institutes of Health (NIH) grant to K.O. (grant GM074207) and by funds from the Hilton Ludwig Cancer Prevention Initiative to A.K.S. F.M. was supported by the German Science Foundation (grant ME 4713/1-1). M.O. and S.W. were supported by the Japan Society for the Promotion of Science. R.L.D, J.V.A, D.A.J., A.K.S., A.D. and K.O. received salary and other support from the Ludwig Institute for Cancer Research. We thank P. Zage for the KPNYN cells.

Author contributions F.M., A.D. and K.O. conceived and designed the study and wrote the manuscript with the support of A.K.S.; F.M. performed all experiments and analysed data unless otherwise noted; M.O. and K.-Y.L. performed all of the coexpression and interaction analyses; R.L.D. and A.K.S. made the initial observation of the sensitivity of neuroblastoma cell lines to centrinone and helped to design and execute tumour xenograft experiments; A.K.S. designed the TRIM37 TRAF-domain mutant; S.W. analysed the coalescence of pericentriolar material; J.V.A. and R.K. helped to analyse the proliferation of the neuroblastoma cell line; D.A.J. helped to analyse gene copy numbers; M.O., K-Y.L., S.W. and R.L.D. contributed to editing of the manuscript.

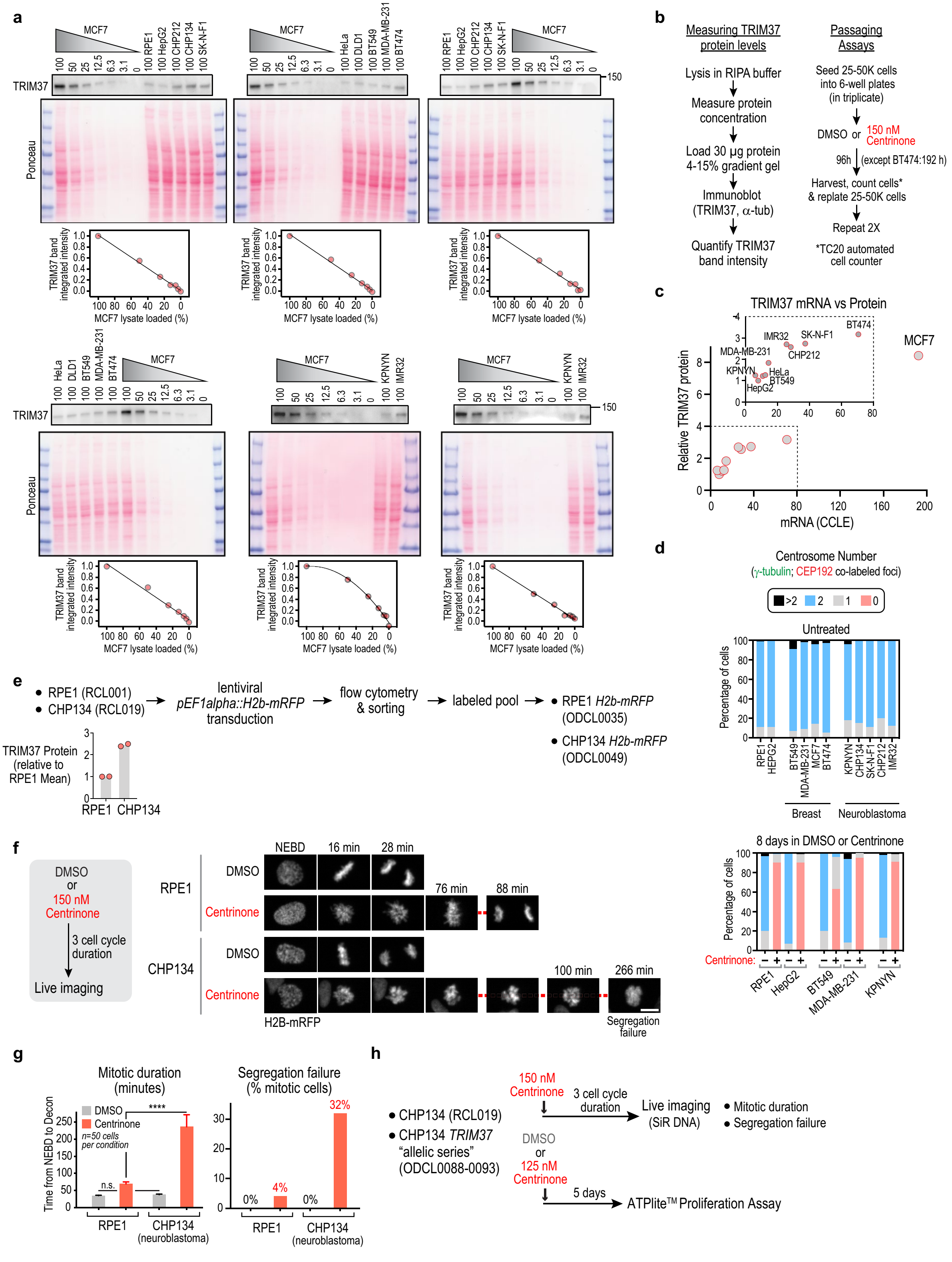
Competing interests K.O., A.K.S., A.D., F.M. and R.L.D. are inventors on a pending patent application from the Ludwig Institute for Cancer Research, PCT/US2018/064243: 'Methods of treating cancer with PLK4 inhibitors'. A.K.S. is also an inventor on a patent application from the Ludwig Institute for Cancer Research, application number PCT/IB2016/000594: 'Plk4 inhibitors'.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2710-1.
Correspondence and requests for materials should be addressed to F.M., A.D. and K.O.
Peer review information *Nature* thanks Renata Basto, Sarah McClelland and William Weiss for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



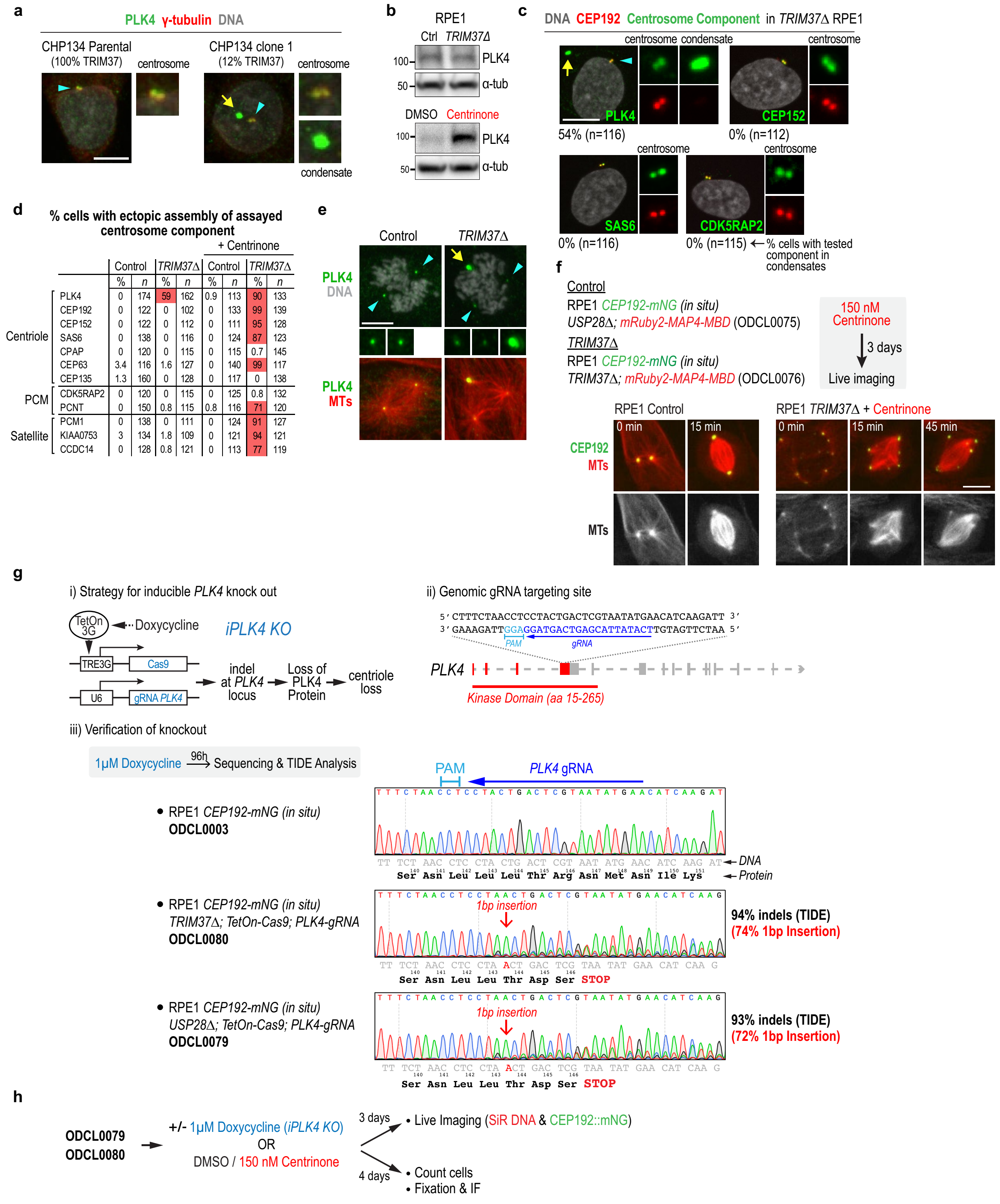
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Effect of varying TRIM37 levels on sensitivity to PLK4 inhibition, and TRIM37 expression profile in patient-derived tumours and cancer cell lines. a, Unique cell line identifiers are used to describe the cell lines in each experiment. The RCL prefix refers to cell lines received from an external source, such as the ATCC. The ODCL prefix refers to cell lines engineered in the Oegema and Desai labs (OD) from received cell lines. **b**, Top, cell-line code and bottom, experimental protocol for the analysis of mitotic duration and chromosome-segregation failure. Clone 5 is the *Tet-ON–TRIM37* cell line shown in Fig. 1a, b, which overexpresses TRIM37 roughly fourfold relative to parental RPE1 cells. **c**, Immunoblots of the RPE1 cell lines described in **b**; transgene-encoded *TetON–TRIM37* expression was induced for 24 h. α-tub, α-tubulin. **d**, **e**, Graphs plotting mitotic duration (**d**) and the frequency of chromosome-segregation failure (**e**) following treatment with DMSO (−) versus centrinone (+) for the three analysed cell lines shown in Fig. 1a, b; $n = 50$ for each condition. Error bars represent 95% confidence intervals. **f**, Live-imaging-based analysis was used to measure mitotic duration and segregation failure for the cell lines described in **b**; values are plotted versus TRIM37 protein level measured by semiquantitative western blotting. Each cell line was treated with DMSO (grey) or centrinone (red) and doxycycline before live imaging; the experimental scheme is shown in **b**. Fifty cells were analysed per condition. Error bars represent 95% confidence intervals. In DMSO, the analysed cell lines exhibited normal mitotic duration and segregation fidelity regardless of TRIM37 protein level. By contrast, in centrinone, loss of TRIM37 reduced mitotic duration and the percentage of cells experiencing segregation failure (green shading), whereas increased TRIM37 protein levels led to a proportional increase in mitotic duration and segregation-failure rate (red shading). **g**, Left, graph plotting *TRIM37* mRNA levels in 2,120 paediatric tumours representing 13 different cancer types (data are from the St Jude PeCan Data Portal[15]). All paediatric cancer types with more than ten tumours analysed are shown. Values for individual tumours (dots) and median values (black lines) are plotted. The three-letter codes to the right of the graph describe the 13 paediatric cancer types. Neuroblastoma (NBL) tumours exhibit the highest TRIM37 expression. Right, box-and-whiskers plot, comparing *TRIM37* mRNA levels in neuroblastoma tumours to those in all other paediatric cancer type tumours. The range represents the 10th to 90th percentiles of the data; the *P*-value shown is from an unpaired *t*-test. $^{****}P < 0.0001$. FPKM, fragments per kilobase of transcript per million mapped reads. **h**, Graph plotting *TRIM37* mRNA levels across cancer cell lines described in the Cancer Cell Line Encyclopedia (CCLE[16]; https://portals.broadinstitute.org/ccle). **i**, mRNA expression versus copy number from CCLE data for breast cancer cell lines. Two cell lines with high *TRIM37* copy number and expression (MCF7 and BT474; red), as well as two cell lines with normal copy number and expression (MDA-MB-231 and BT549; green) are marked. **j**, List of breast cancer and neuroblastoma cell lines used for analysis in Fig. 1d, e. HepG2 is a hepatocellular carcinoma derived cell line with similar TRIM37 expression to control RPE1 cells. For gel source data see Supplementary Fig. 1. For details on statistics, see Methods; unpaired *t*-tests assuming equal standard deviation were performed.



Extended Data Fig. 2 | See next page for caption.

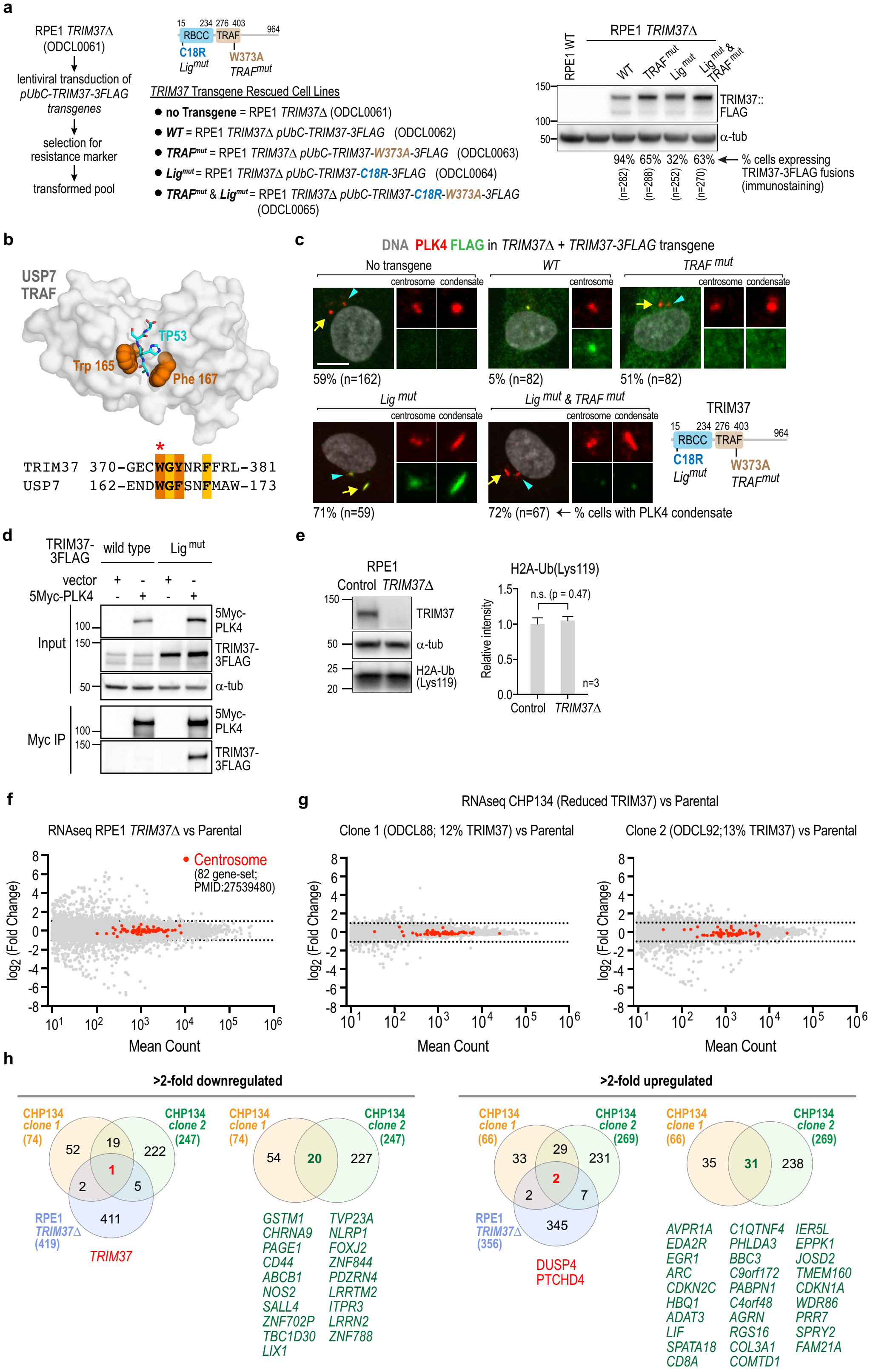
Extended Data Fig. 2 | Analysis of TRIM37 protein levels and centrinone efficacy in different cancer cell lines, and comparison of mitosis in RPE1 and CHP134 neuroblastoma cells following centrinone treatment. **a**, Immunoblots used to quantify TRIM37 protein levels across different cell lines. TRIM37 immunoblots are shown above the corresponding Ponceau-stained blots. MCF7 cells have the highest *TRIM37* transcript levels and copy number in the CCLE (Extended Data Fig. 1i). Serial dilutions of MCF7 extracts were loaded next to extracts from other cell lines on each blot, and TRIM37 band intensities across a serial dilution of MCF7 cell extract were used to generate a standard curve (graphs below each blot); measured intensities for other cell line extracts were converted into relative expression values using the standard curve. TRIM37 protein level in HepG2 cells was set to 1 and values measured for other cell lines were plotted relative to the HepG2 level in Fig. 1d. **b**, Protocols used to measure TRIM37 protein levels (left) and conduct passaging-based proliferation analysis of cancer cell lines (right). **c**, Comparison of TRIM37 mRNA and protein levels across cancer cell lines. mRNA levels are from the CCLE and were transformed from a logarithmic (base 2) to a linear scale. Protein levels are mean values from two measurements, taken as in **a**, and are plotted relative to the amount of TRIM37 in HepG2, a non-amplified cancer cell line. The inset graph excludes MCF7, which shows exceptionally high TRIM37 mRNA and protein levels. **d**, Quantification of centrosome number in the indicated cell lines and conditions ($n = 100$ for each condition). Centrosomes were defined as co-localized foci of CEP192 and γ-tubulin in fixed interphase cells. In the absence of any treatments, there is mild centrosome amplification in the breast cancer cell lines and in one neuroblastoma cell line. Following an eight-day treatment with centrinone, a substantial proportion of the cells from cell lines with relatively low sensitivity to centrinone lacked centrosomes. **e**, Method used to generate a pool of cells expressing H2b–mRFP for the indicated cell lines. TRIM37 protein levels are shown relative to levels in RPE1 cells, measured by semiquantitative immunoblotting. **f**, Images are stills from time-lapse sequences of H2b–RFP-expressing mitotic RPE1 and CHP134 cells. Both cell lines exhibit rapid mitosis (taking around 30 min) with no segregation failure in DMSO. Following centrinone treatment, CHP134 cells exhibit more delayed mitosis and higher rates of segregation failure compared with RPE1 cells. Scale bar, 10 μm. **g**, Quantification of mitotic duration and segregation failure, comparing RPE1 and CHP134 cells. **h**, Protocols used to analyse mitotic duration, segregation failure and viability of the CHP134-derived cell lines with different levels of TRIM37 protein. For details on statistics, see Methods; unpaired *t*-tests assuming equal standard deviation were performed.



		Control		*TRIM37Δ*		+ Centrinone Control		+ Centrinone *TRIM37Δ*	
		%	*n*	%	*n*	%	*n*	%	*n*
Centriole	PLK4	0	174	59	162	0.9	113	90	133
	CEP192	0	122	0	102	0	133	99	139
	CEP152	0	122	0	112	0	111	95	128
	SAS6	0	138	0	116	0	124	87	123
	CPAP	0	120	0	115	0	115	0.7	145
	CEP63	3.4	116	1.6	127	0	140	99	117
	CEP135	1.3	160	0	128	0	117	0	138
PCM	CDK5RAP2	0	120	0	115	0	125	0.8	132
	PCNT	0	150	0.8	115	0.8	116	71	120
Satellite	PCM1	0	138	0	111	0	124	91	127
	KIAA0753	3	134	1.8	109	0	121	94	121
	CCDC14	0	128	0.8	121	0	113	77	119

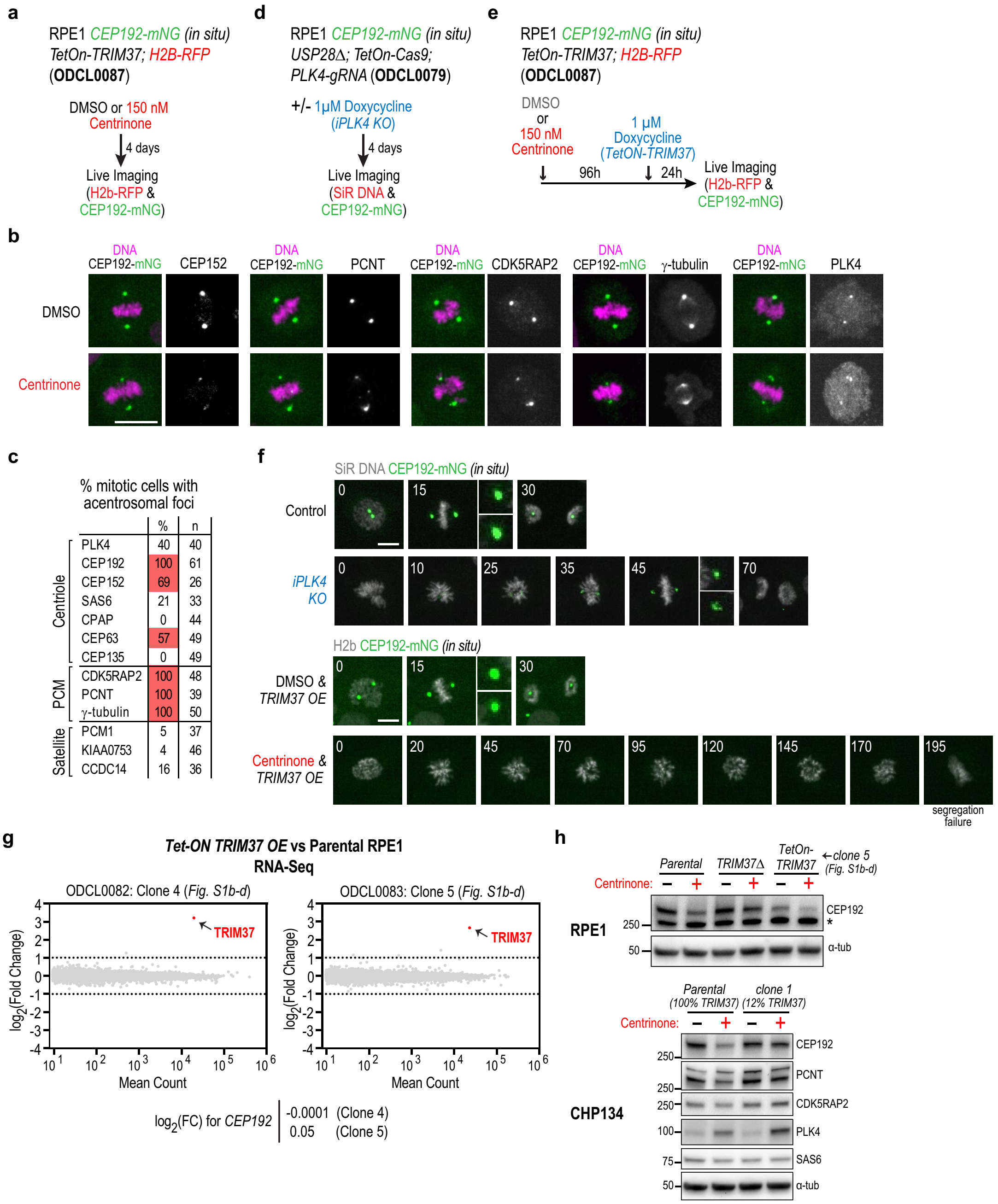
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Analysis of *TRIM37Δ* cells, rescue with *TRIM37* variants and generation of the inducible *PLK4* knockout. a, Formation of PLK4 condensates in CHP134 neuroblastoma cells with reduced TRIM37 expression. Parental CHP134 cells, which have four copies of the *TRIM37* gene, were compared with clone 1 (Fig. 1f; roughly 12% TRIM37 expression relative to parental CHP134). PLK4 condensates were observed in 23% of the clone 1 cells with reduced TRIM37 expression but in none of the parental cells ($n = 100$ for each). **b**, Immunoblots of RPE1 cells, comparing the effect on PLK4 protein levels of *TRIM37* deletion (top) versus inhibition of PLK4 kinase activity using centrinone (bottom). PLK4 protein levels were elevated roughly sevenfold following the inhibition of kinase activity (7.4 ± 1.1 fold; mean ± s.d.; $n = 3$), confirming that the detected band corresponds to PLK4. The *TRIM37Δ* blot is the same as in Fig. 2b. **c**, Immunofluorescence images of the indicated centrosomal components in *TRIM37Δ* cells. Scale bar, 10 µm. **d**, Summary of immunofluorescence analysis in interphase cells. PCM, pericentriolar material. **e**, Immunofluorescence image showing microtubule organization by a PLK4 condensate in a mitotic *TRIM37Δ* cell. Scale bar, 10 µm. **f**, Top, protocol used to conduct live imaging of CEP192 and microtubules. Bottom, images of control and centrinone-treated *TRIM37Δ* cells with in situ mNG-tagged CEP192 and a transgene that expresses a red fluorescent microtubule-binding domain (MBD). Times in minutes after NEBD are noted on each panel. Scale bar, 10 µm. The merged *TRIM37Δ* images are the same as those shown in Fig. 2e. **g**, Description and validation of the inducible *PLK4* knockout engineered in *TRIM37Δ* and control (*USP28Δ*) cells. *USP28Δ* cells were used as the control because inactivation of USP28 prevents the p53 activation and G1 arrest that are observed as a consequence of delayed mitosis following centrosome loss in RPE1 cells[13]. Note that *USP28Δ* has no effect on the mitotic consequences of centrosome loss[13] and enables comparison with *TRIM37Δ* cells, which prevent p53 activation following centrinone treatment by accelerating mitosis in the absence of centrosomes. The gRNA sequence used to target *PLK4* exon 5 is depicted, and the efficacy of the inducible knockout in both cell lines was validated by inducing Cas9 expression using doxycycline for four days, followed by sequencing and tracking of indels by decomposition (TIDE) analysis[35]. Sequence traces show a high frequency of indels, with a 1-bp insertion being the most frequent outcome. **h**, Protocol used to compare centrinone treatment with the *iPLK4 KO* in Fig. 2f, g. For gel source data, see Supplementary Fig. 1.

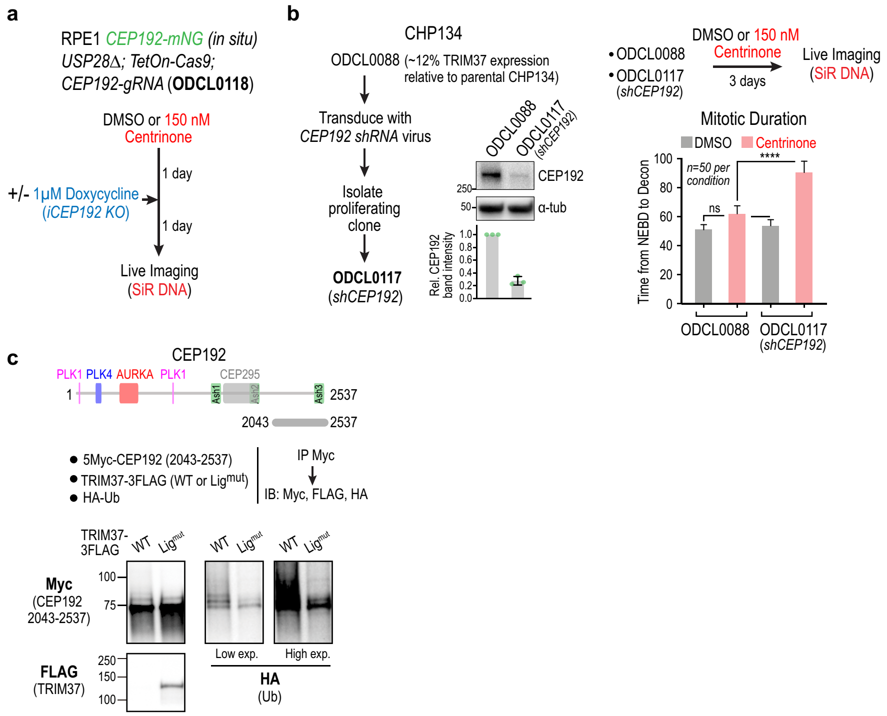


Extended Data Fig. 4 | See next page for caption.

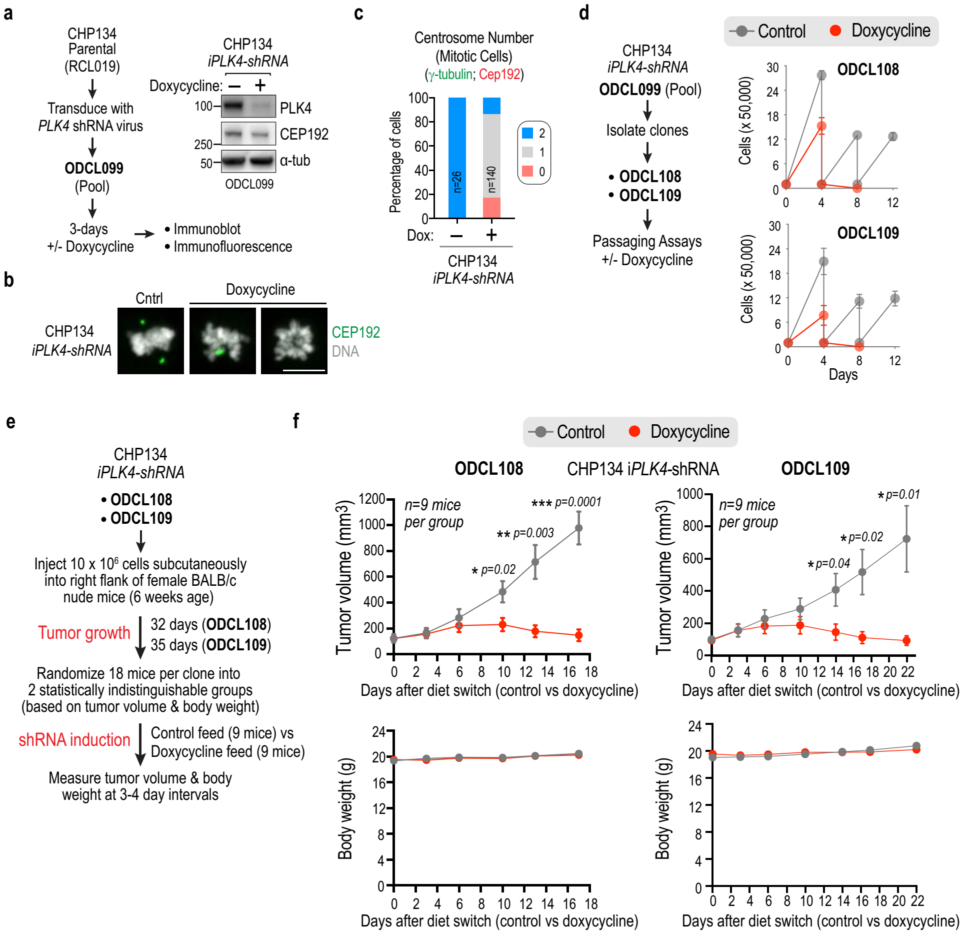
Extended Data Fig. 4 | Generation of TRIM37 variants, and analysis of the effect of TRIM37 loss on the ubiquitination of histone H2A and transcription. **a**, Left, method for generating the cell lines used to test rescue with transgenes encoding wild-type and mutant TRIM37. Centre, top, the point mutations engineered in the ligase and TRAF domains of TRIM37. Right, blot showing the expression of transgene-encoded TRIM37 variants in the pools selected for marker resistance; the percentages of cells expressing the indicated fusions are shown below the blot. **b**, Structural/sequence analysis used to engineer the TRIM37 TRAF-domain mutant. The USP7 TRAF domain (grey surface) is shown bound to a p53 peptide (cyan stick), with key binding residues W165 and F167 in orange spheres (Protein DataBank (https://www.rcsb.org) code 3MQR). The sequences show the similarity between the peptide-binding pockets of TRIM37 and USP7; the conserved tryptophan (W165 in USP7; W373 in TRIM37) was mutated to alanine to generate the TRIM37 TRAF mutant. **c**, Images illustrating the effect of expressing wild-type (WT) TRIM37 or engineered variants disrupting ligase activity or TRAF-domain interactions in *TRIM37Δ* cells. **d**, Interaction analysis using coexpression of TRIM37 and PLK4 followed by immunoprecipitation of PLK4. WT TRIM37 is expressed at substantially lower levels than ligase-mutant (C18R) TRIM37, suggesting that TRIM37 autoregulates its own stability. The low expression of WT TRIM37 led us to use ligase-mutant TRIM37 for the interaction analysis shown in Fig. 2i. **e**, Left, immunoblot of H2A conjugated via lysine 119 to ubiquitin (Ub), comparing control and *TRIM37Δ* RPE1 cells. α-Tubulin (α-tub) served as a loading control. Right, quantification of band intensities indicates that TRIM37 does not reduce ubiquitination of Lys 119 in histone H2A. **f**, RNA-sequencing (RNA-seq) analysis comparing parental and *TRIM37Δ* RPE1 cells. A previously defined set of 82 genes encoding centrosomal components[43] is marked in red to highlight the lack of change in their mRNA levels. PMID, PubMed identification code. **g**, RNA-seq analysis comparing parental CHP134 cells with two clones (clones 1 and 2 from Fig. 1f) with substantially lower expression of TRIM37. The centrosome 82-gene set is highlighted in red. **h**, Lists of genes that are more than twofold downregulated or upregulated. Each of the three test lines (RPE1 *TRIM37Δ*, CHP134 clone 1 and CHP134 clone 2) was compared with the parental line in order to identify genes with statistically significant, more than twofold changes. Cross-comparison of all three test lines and of the two CHP134 clones is summarized in the Venn diagrams; gene names for shared differentially expressed genes are shown below each Venn diagram. For gel source data, see Supplementary Fig. 1. For details on statistics, see Methods; unpaired *t*-tests assuming equal standard deviation were performed.



Extended Data Fig. 5 | Effect of TRIM37 overexpression on the coalescence of pericentriolar material and acentrosomal mitosis. **a**, Protocol used to analyse the effect of centrinone treatment on the mitotic dynamics of the CEP192 protein by live imaging (Fig. 3a). **b**, Immunofluorescence images of centrosome components in DMSO- versus centrinone-treated mitotic RPE1 cells. Scale bar, 10 µm. **c**, Summary of immunofluorescence analysis, showing which centrosome components were detected in the foci at the poles of acentrosomal spindles in centrinone-treated cells. Scale bar, 10 µm. **d**, Protocol used to inducibly knock out *PLK4* and monitor CEP192 dynamics in mitosis by live imaging (Fig. 3b). **e**, Protocol used to overexpress TRIM37 and monitor CEP192 dynamics in mitosis by live imaging (Fig. 3c). Note that this is the same cell line used for the analysis in Fig. 3a (no doxycline induction); the analyses in these two conditions were conducted in parallel. **f**, Additional panels from the time-lapse image sequences shown in Fig. 3b, c for inducible *PLK4* knockout and TRIM37 overexpression. Times in minutes after NEBD are noted on each panel. **g**, RNA-seq analysis comparing two clones that overexpressed TRIM37 with parental RPE1 cells. Elevated *TRIM37* transcript levels are evident in both clones. No significant changes in the global transcriptome were otherwise observed. **h**, Evidence that centrosomes protect CEP192 from TRIM37-dependent degradation. Top, immunoblots of the indicated RPE1 cell lines with and without centrinone treatment. In cells overexpressing TRIM37, centrinone treatment further reduces CEP192 levels; by contrast, in *TRIM37*Δ cells, centrinone treatment affects CEP192 levels only modestly. Note that RNA-seq analysis indicated no significant change in CEP192 transcript levels between cell lines with varying levels of TRIM37. Bottom, immunoblots of CHP134 parental cells and a derived clone with TRIM37 expression levels roughly 12% of those of the parental cells. Centrinone strongly reduced CEP192 levels in the parental cells but not in the clone with reduced TRIM37 expression. α-Tubulin serves as a loading control. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 6 | Evidence that CEP192 is the target of TRIM37 that accounts for enhanced sensitivity to PLK4 inhibition. a, Protocol for partial CEP192 inhibition using a short-term inducible knockout, followed by live imaging of mitosis. **b**, Evidence in CHP134 cells that CEP192 is a functionally important target of TRIM37. A CHP134 clonal cell line with reduced TRIM37 expression (roughly 12% relative to parental CHP134 cells) was stably transduced with a *CEP192* shRNA that reduced expression by approximately 75% (immunoblot and quantification below). Right, live imaging of mitosis showed that although reduction of CEP192 levels had no significant effect on the duration of mitosis in DMSO-treated cells, it did significantly extend mitotic duration in centrinone-treated cells. $****P < 0.0001$. **c**, Evidence that the C terminus of CEP192 is ubiquitinated in a TRIM37-dependent manner. The experiment shown in Fig. 4d included co-transfection of HA-tagged ubiquitin. Shown here is the HA–ubiquitin blot (together with FLAG and Myc blots) of the immunoprecipitated C-terminal CEP192 fragment that binds TRIM37. Ubiquitination of this fragment was enhanced in the presence of WT relative to ligase-mutant TRIM37. The FLAG blot shown is the same as in Fig. 4d; the Myc blot is a different exposure of that in Fig. 4d. The other CEP192 fragments are not shown because their stability was affected by coexpression with WT but not ligase-mutant TRIM37, which makes comparisons of ubiquitination profiles difficult. For gel source data, see Supplementary Fig. 1. For details on statistics, see Methods; unpaired *t*-tests assuming equal standard deviation were performed.



Extended Data Fig. 7 | Analysis of CHP134 xenograft tumours on the basis of inducible *PLK4* shRNA. **a**, Left, generation and characterization of a CHP134 pool with stably integrated inducible *PLK4* (*iPLK4*) shRNA. Following viral transduction, three-day induction of the shRNA with doxycycline followed by immunoblotting and immunofluorescence was used to assess PLK4 depletion and centrosome loss. Right, the immunoblot shows depletion of PLK4 as well as reduction of CEP192 following doxycycline treatment, as is also observed with centrinone; this reduction depended on high TRIM37 expression in CHP134 cells, as it was not observed following induction of *PLK4* shRNA in a CHP134-derived line with roughly 10% TRIM37 expression (data not shown). **b**, Immunofluorescence images showing loss of centrosomes, detected using CEP192, following induction of *PLK4* shRNA. Scale bar, 10 µm. **c**, Quantification of centrosome number after three-day induction of *PLK4* shRNA. A longer induction was associated with extensive lethality, as is also observed with centrinone treatment of CHP134 cells. **d**, Protocol for isolating CHP134 clones with stably integrated *iPLK4* shRNA from the pool described in **a**–**c**. Right, results of passaging-based analysis, showing that both clones exhibited rapid cessation of proliferation following induction of the shRNA. **e**, Workflow for tumour xenograft analysis with the two CHP134 *iPLK4*-shRNA clones. Tumours were generated in female BALB/c nude mice, and the shRNA was induced by switching to a doxycycline-containing diet. Tumour volume and body weight were measured over time after induction. **f**, Time course of tumour growth in BALB/c nude mice for the two CHP134 *PLK4* shRNA clonal lines following induction of shRNA (doxycycline) versus no induction (control). Error bars show standard error of the mean. Statistical significance was evaluated using unpaired *t*-tests (Welch's correction). For gel source data, see Supplementary Fig. 1.

Extended Data Table 1 | Human cell lines used

Parental cell lines

OD cell line code	Name	Source	Clonal or Polyclonal	Catalog #
RCL001	hTERT RPE-1	ATCC	n/a	CRL-4000
RCL008	CHP212	ATCC	n/a	CRL-2273
RCL010	IMR32	ATCC	n/a	CCL-127
RCL019	CHP134	Sigma	n/a	06122002
RCL020	SK-N-F1	ATCC	n/a	CRL-2142
RCL021	HepG2	ATCC	n/a	HB-8065
RCL022	MCF7	ATCC	n/a	HTB-22
RCL023	BT-474	ATCC	n/a	HTB-20
RCL024	BT-549	ATCC	n/a	HTB-122
RCL025	MDA-MB-231	ATCC	n/a	HTB-26
RCL026	Freestyle 293-F	Thermo Fisher Scientific	n/a	R79007
RCL027	KPNYN	Gift from Peter Zage (UCSD)		n/a

Engineered cell lines

OD cell line code	Parental line	Modification(s)	Clonal or Polyclonal	Reference
ODCL0002	hTERT RPE-1	*USP28Δ*	Clonal	13
ODCL0003	hTERT RPE-1	*CEP192-mNeonGreen*	Clonal	This study
ODCL0035	hTERT RPE-1	*EF-1*α^{pro}*-H2B-mRFP*	Polyclonal	13
ODCL0036	CHP212	*EF-1*α^{pro}*-H2B-mRFP*	Polyclonal	This study
ODCL0049	CHP134	*EF-1*α^{pro}*-H2B-mRFP*	Polyclonal	This study
ODCL0060	SK-N-F1	*EF-1*α^{pro}*-H2B-mRFP*	Polyclonal	This study
ODCL0061	hTERT RPE-1	*TRIM37Δ*	Clonal	13
ODCL0062	ODCL0061	*TRIM37Δ UbC*^{pro}*-TRIM37-3xFLAG*	Polyclonal	This study
ODCL0063	ODCL0061	*TRIM37Δ UbC*^{pro}*-TRIM37-W373A-3xFLAG*	Polyclonal	This study
ODCL0064	ODCL0061	*TRIM37Δ UbC*^{pro}*-TRIM37-C18R-3xFLAG*	Polyclonal	This study
ODCL0065	ODCL0061	*TRIM37Δ UbC*^{pro}*-TRIM37-C18R-W373A-3xFLAG*	Polyclonal	This study
ODCL0068	ODCL0061	*TRIM37Δ PGK*^{pro}*-TRIM37-C18R-mNeonGreen*	Polyclonal	This study
ODCL0070	ODCL0068	*TRIM37Δ PGK*^{pro}*-TRIM37-C18R-mNeonGreen* *EF-1*α^{pro}*-mRuby2-MAP4-MBD*	Polyclonal	This study
ODCL0071	ODCL0003	*CEP192-mNeonGreen; USP28Δ*	Clonal	This study
ODCL0072	ODCL0071	*CEP192-mNeonGreen; USP28Δ; EF-1*α^{pro}*-H2B-mRFP*	Polyclonal	This study
ODCL0073	ODCL0003	*CEP192-mNeonGreen; TRIM37Δ*	Clonal	This study
ODCL0074	ODCL0073	*CEP192-mNeonGreen; TRIM37Δ; EF-1*α^{pro}*-H2B-mRFP*	Polyclonal	This study
ODCL0075	ODCL0071	*CEP192-mNeonGreen; USP28Δ; EF-1*α^{pro}*-mRuby2-MAP4-MBD*	Polyclonal	This study
ODCL0076	ODCL0073	*CEP192-mNeonGreen; TRIM37Δ; EF-1*α^{pro}*-mRuby2-MAP4-MBD*	Polyclonal	This study
ODCL0077	ODCL0071	*CEP192-mNeonGreen; USP28Δ; TRE3G*^{pro}*-Cas9*	Clonal	This study
ODCL0078	ODCL0073	*CEP192-mNeonGreen; TRIM37Δ; TRE3G*^{pro}*-Cas9*	Clonal	This study
ODCL0079	ODCL0077	*CEP192-mNeonGreen; USP28Δ; TRE3G*^{pro}*-Cas9; U6*^{pro}*-gRNA-PLK4*	Polyclonal	This study
ODCL0080	ODCL0078	*CEP192-mNeonGreen; TRIM37Δ; TRE3G*^{pro}*-Cas9; U6*^{pro}*-PLK4-gRNA*	Polyclonal	This study
ODCL0081	RCL001	*TRE3GS*^{pro}*-TRIM37*	Clone 3; ED Fig. 1b-f	This study
ODCL0082	RCL001	*TRE3GS*^{pro}*-TRIM37*	Clone 4; Fig. 3g, ED Fig. 1b-f	This study
ODCL0083	RCL001	*TRE3GS*^{pro}*-TRIM37*	Clone 5; Fig. 1a-b, 3g, ED Fig. 1b-f	This study
ODCL0084	ODCL0061	*TRIM37Δ; TRE3GS*^{pro}*-TRIM37*	Clone 1; ED Fig. 1b-f	This study
ODCL0085	ODCL0061	*TRIM37Δ; TRE3GS*^{pro}*-TRIM37*	Clone 2; ED Fig. 1b-f	This study
ODCL0086	ODCL0003	*CEP192-mNeonGreen; TRE3GS*^{pro}*-TRIM37*	Clonal	This study
ODCL0087	ODCL0086	*CEP192-mNeonGreen; TRE3GS*^{pro}*-TRIM37; EF-1*α^{pro}*-H2B-mRFP*	Polyclonal	This study
ODCL0088	CHP134	*TRIM37+/-/-/- (17bp del; 7bp del; 1bp ins)*	Clone 1; Fig. 1f, 3f, ED Fig. 2h	This study
ODCL0089	CHP134	*TRIM37+/+/+/- (374bp del)*	Clone 6; Fig. 1f, ED Fig. 2h	This study
ODCL0090	CHP134	*TRIM37+/-/-/- (15bp del; 1bp ins; 7bp ins)*	Clone 4; Fig. 1f, ED Fig. 2h	This study
ODCL0091	CHP134	*TRIM37+/+/+/- (2bp del)*	Clone 5; Fig. 1f, ED Fig. 2f	This study
ODCL0092	CHP134	*TRIM37+/-/-/- (1bp del; 1bp del; 17bp ins)*	Clone 2; Fig. 1f, ED Fig. 2h	This study
ODCL0093	CHP134	*TRIM37+/+/+/- (3 copies wildtype; mutation unclear)*	Clone 3; Fig. 1f, ED Fig. 2h	This study
ODCL0099	CHP134	*hEF1a/TurboGFP inducible PLK4-shRNA*	Polyclonal	This study
ODCL0108	ODCL0099	*hEF1a/TurboGFP inducible PLK4-shRNA*	Clone 1; ED Fig 7d-f	This study
ODCL0109	ODCL0099	*hEF1a/TurboGFP inducible PLK4-shRNA*	Clone 2; ED Fig 7d-f	This study
ODCL0117	ODCL0088	*TRIM37+/-/-/- (clone 1); CEP192-shRNA*	Clonal	This study
ODCL0118	hTERT RPE-1	*CEP192-mNeonGreen; USP28Δ;* *TRE3G*^{pro}*-Cas9; U6*^{pro}*-gRNA-CEP192*	Polyclonal	This study
ODCL0119	ODCL0061	*TRIM37Δ; UBC*^{pro}*-TRIM37*	Clonal	This study
ODCL0121	ODCL0061	*TRIM37Δ; UBC*^{pro}*-TRIM37-C18R*	Clonal	This study

The RCL prefix refers to cell lines received from an external source, such as the ATCC. The ODCL prefix refers to cell lines engineered in the Oegema and Desai labs (OD) from received cell lines. Del, deletion; ins, insertion; pro, promoter.

Extended Data Table 2 | Plasmids used

OD Plasmid code	Description	Purpose	Selection	Reference
pOD3789	CMV^{pro}-5xMYC-PLK4	Transient transfection	Ampicillin	This study
pOD3790	CMV^{pro}-TRIM37-3xFLAG	Transient transfection	Ampicillin	This study
pOD3791	CMV^{pro}-TRIM37-C18R-3xFLAG	Transient transfection	Ampicillin	This study
pOD3792	CMV^{pro}-TRIM37-C18R-W373A-3xFLAG	Transient transfection	Ampicillin	This study
pOD3793	CMV^{pro}-5xMYC-CEP152	Transient transfection	Ampicillin	This study
pOD3794	CMV^{pro}-5xMYC-CEP192	Transient transfection	Ampicillin	This study
pOD3795	EF-1α^{pro}-H2B-mRFP	Lentiviral integration	Ampicillin	13
pOD3796	EF-1α^{pro}-mRuby-hMAP4-MBP; SV40pro-NeoR	Lentiviral integration	Ampicillin	This study
pOD3797	UbC^{pro}-TRIM37-3xFLAG SV40pro-NeoR	Lentiviral integration	Ampicillin	This study
pOD3798	UbC^{pro}-TRIM37-C18R-3xFLAG; SV40pro-NeoR	Lentiviral integration	Ampicillin	This study
pOD3799	UbC^{pro}-TRIM37-C18R-W373A-3xFLAG; SV40pro-NeoR	Lentiviral integration	Ampicillin	This study
pOD3800	UbC^{pro}-TRIM37-C18R-W373A-3xFLAG; SV40pro-NeoR	Lentiviral integration	Ampicillin	This study
pOD3801	PGK^{pro}-TRIM37-C18R-NeonGreen-P2A-BSD	Lentiviral integration	Ampicillin	This study
pOD3802	U6^{pro}-TRIM37gRNA (PX459)	Transient transfection	Ampicillin	13
pOD3803	U6^{pro}-USP28-gRNA (PX459)	Transient transfection	Ampicillin	13
pOD3804	U6^{pro}-PLK4-gRNA (lentiGuide-Puro)	Lentiviral integration	Ampicillin	This study
pOD3805	TRE3GS^{pro}-TRIM37 SV40pro-NeoR	Lentiviral integration	Ampicillin	This study
pOD3806	PGK^{pro}-Tet-On-3G-P2A-BSD	Lentiviral integration	Ampicillin	This study
pOD3807	UbC^{pro}-TRIM37 SV40pro-NeoR	Lentiviral integration	Ampicillin	This study
pOD3808	UbC^{pro}-TRIM37-C18R SV40pro-NeoR	Lentiviral integration	Ampicillin	This study
pOD3810	CMV^{pro}-5xMYC-Cep192 aa 1-2071	Transient transfection	Ampicillin	This study
pOD3811	CMV^{pro}-5xMYC-Cep192 aa 1201-2537	Transient transfection	Ampicillin	This study
pOD3812	CMV^{pro}-5xMYC-Cep192 aa 2043-2537	Transient transfection	Ampicillin	This study

Pro, promoter.

nature research

Corresponding author(s): Karen Oegema
Last updated by author(s): May 4, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Only commercial software was used. In particular: CQ1 Software (Yokogawa); Cell Voyager Measurement System R2.02.07 (Yokogawa); Image Lab 4.1 (BioRad)
Data analysis	Only commercial software was used. In particular: FIJI, Excel, Prism, BaseSpace

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All image source data is available upon request, RNA-Seq data is being submitted to GEO (accession number GSE148263. There is associated raw data that has been uploaded for Extended Data Figure 7f.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes are indicated in all of the figure legends. Sample sizes were always at or above the standard for this type of analysis and only differences well above the level of statistical significance are highlighted.
Data exclusions	No data was excluded.
Replication	All growth curves were done in triplicate. Quantitative western blotting was done twice. In addition the majority of experiments were performed in two cell type backgrounds (RPE1 and CHP134), with similar results being obtained in each. In cases where clonal cell lines were used, more than one independent clone was analyzed unless otherwise noted.
Randomization	For the mouse xenograft experiments in Extended Data Figure 7f, the mice were randomly allocated into test groups, while ensuring comparable mean initial tumor size, by BIODURO, the company that performed the study.
Blinding	Blinding was not used for any of the reported experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	All antibodies, with their vendor, catalog number, and working concentrations, are described in the methods section.
Validation	The antibodies to the key proteins TRIM37, PLK4 and CEP192 were validated as described. TRIM37 antibodies were validated for western blotting by comparing extracts of TRIM37 deleted or overexpressing cells to unperturbed control cells. PLK4 antibodies were validated using western blots of cells before and after centrinone treatment, which leads to a substantial increase in PLK4 levels. CEP192 antibodies were validated by blotting cells following CEP192 knockdown using an shRNA and also using an inducible CRISPR knockout cell line. For antibodies to other centrosomal components, we used antibodies that were cited in multiple other studies and confirmed that all recognize a band of the correct size on western blots and centrosomes by immunofluorescence in control cells.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The majority of parental cell lines were obtained from the ATCC (American Type Culture Collection), except CHP134, which was obtained from Sigma, the Freestyle 293-F cells used for co-expression, which were obtained from Thermo Fisher Scientific and the KPNYN cells, which were a gift from Peter Zage.
Authentication	The majority of cell lines were obtained directly from the ATCC and were not revalidated.

Mycoplasma contamination	All cell lines were tested every 3 months for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	For the experiment in Extended Data Figure 7f, 4 groups of 9 female BALB/c mice were used. This experiment was outsourced and performed by BIODURO. All the procedures related to animal handling, care and the treatment in this study were performed according to the guidelines approved by the Institutional Animal Care and Use Committee (IACUC) of BioDuro following the guidance of the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC).
Wild animals	n/a
Field-collected samples	n/a
Ethics oversight	All the procedures related to animal handling, care and the treatment in this study were performed according to the guidelines approved by the Institutional Animal Care and Use Committee (IACUC) of BioDuro following the guidance of the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Targeting TRIM37-driven centrosome dysfunction in 17q23-amplified breast cancer

https://doi.org/10.1038/s41586-020-2690-1

Received: 10 June 2019

Accepted: 17 June 2020

Zhong Y. Yeow[1,2,6], Bramwell G. Lambrus[3,6], Rebecca Marlow[4,5], Kevin H. Zhan[3], Mary-Anne Durin[1,2], Lauren T. Evans[3], Phillip M. Scott[3], Thao Phan[3], Elizabeth Park[3], Lorena A. Ruiz[3], Daniela Moralli[2], Eleanor G. Knight[4], Luned M. Badder[5], Daniela Novo[4], Syed Haider[4], Catherine M. Green[2], Andrew N. J. Tutt[4,5], Christopher J. Lord[4], J. Ross Chapman[1,2]✉ & Andrew J. Holland[3]✉

Genomic instability is a hallmark of cancer, and has a central role in the initiation and development of breast cancer[1,2]. The success of poly-ADP ribose polymerase inhibitors in the treatment of breast cancers that are deficient in homologous recombination exemplifies the utility of synthetically lethal genetic interactions in the treatment of breast cancers that are driven by genomic instability[3]. Given that defects in homologous recombination are present in only a subset of breast cancers, there is a need to identify additional driver mechanisms for genomic instability and targeted strategies to exploit these defects in the treatment of cancer. Here we show that centrosome depletion induces synthetic lethality in cancer cells that contain the 17q23 amplicon, a recurrent copy number aberration that defines about 9% of all primary breast cancer tumours and is associated with high levels of genomic instability[4–6]. Specifically, inhibition of polo-like kinase 4 (PLK4) using small molecules leads to centrosome depletion, which triggers mitotic catastrophe in cells that exhibit amplicon-directed overexpression of *TRIM37*. To explain this effect, we identify TRIM37 as a negative regulator of centrosomal pericentriolar material. In 17q23-amplified cells that lack centrosomes, increased levels of TRIM37 block the formation of foci that comprise pericentriolar material—these foci are structures with a microtubule-nucleating capacity that are required for successful cell division in the absence of centrosomes. Finally, we find that the overexpression of TRIM37 causes genomic instability by delaying centrosome maturation and separation at mitotic entry, and thereby increases the frequency of mitotic errors. Collectively, these findings highlight *TRIM37*-dependent genomic instability as a putative driver event in 17q23-amplified breast cancer and provide a rationale for the use of centrosome-targeting therapeutic agents in treating these cancers.

Many cancer cells can proliferate without centrosomes[7,8]. However, while evaluating the response of cell lines to centrosome loss, we discovered that MCF-7 human breast adenocarcinoma cells were hypersensitive to centrosome loss induced by treatment with the PLK4 inhibitor centrinone[7]. Progressive centrosome loss induced upon treatment with centrinone in MCF-7 cells (Extended Data Fig. 1a) blocked the proliferation of these cells within three days (Fig. 1a), and greatly reduced clonogenic survival (Fig. 1c). In non-transformed cells, centrosome depletion using centrinone leads to activation of the mitotic surveillance pathway, which triggers p53-dependent growth arrest through USP28 and 53BP1[7,9–13]. However, we found that the sensitivity of MCF-7 cells to centrinone treatment was independent of this pathway (Fig. 1a, Extended Data Fig. 1b).

We considered whether the genetic background of MCF-7 cells underlies their hypersensitivity to centrosome depletion induced by PLK4 inhibition. MCF-7 cells contain the 17q23 breast cancer amplicon, a 3–4-Mb recurrent copy number aberration found in about 9% of all primary breast cancer tumours[4,14]. The 17q23 amplification also represents the defining feature of IntClust1 tumours, a subset of primarily oestrogen-receptor (ER)-positive, luminal B-type breast cancers that was detected following the genomic and transcriptomic profiling of more than 2,000 primary breast tumours by the METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) project[4,5,15]. Of the approximately 40 protein-coding genes located within the 17q23 amplicon[16–18], we noted the presence of *TRIM37*, a gene that has previously been implicated in centrosome function[11,19].

---

[1]Medical Research Council (MRC) Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. [2]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. [3]Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [4]The Breast Cancer Now Toby Robins Breast Cancer Research Centre, The Institute of Cancer Research, London, UK. [5]The Breast Cancer Now Unit, King's College London, London, UK. [6]These authors contributed equally: Zhong Y. Yeow, Bramwell G. Lambrus. ✉e-mail: ross.chapman@imm.ox.ac.uk; aholland@jhmi.edu

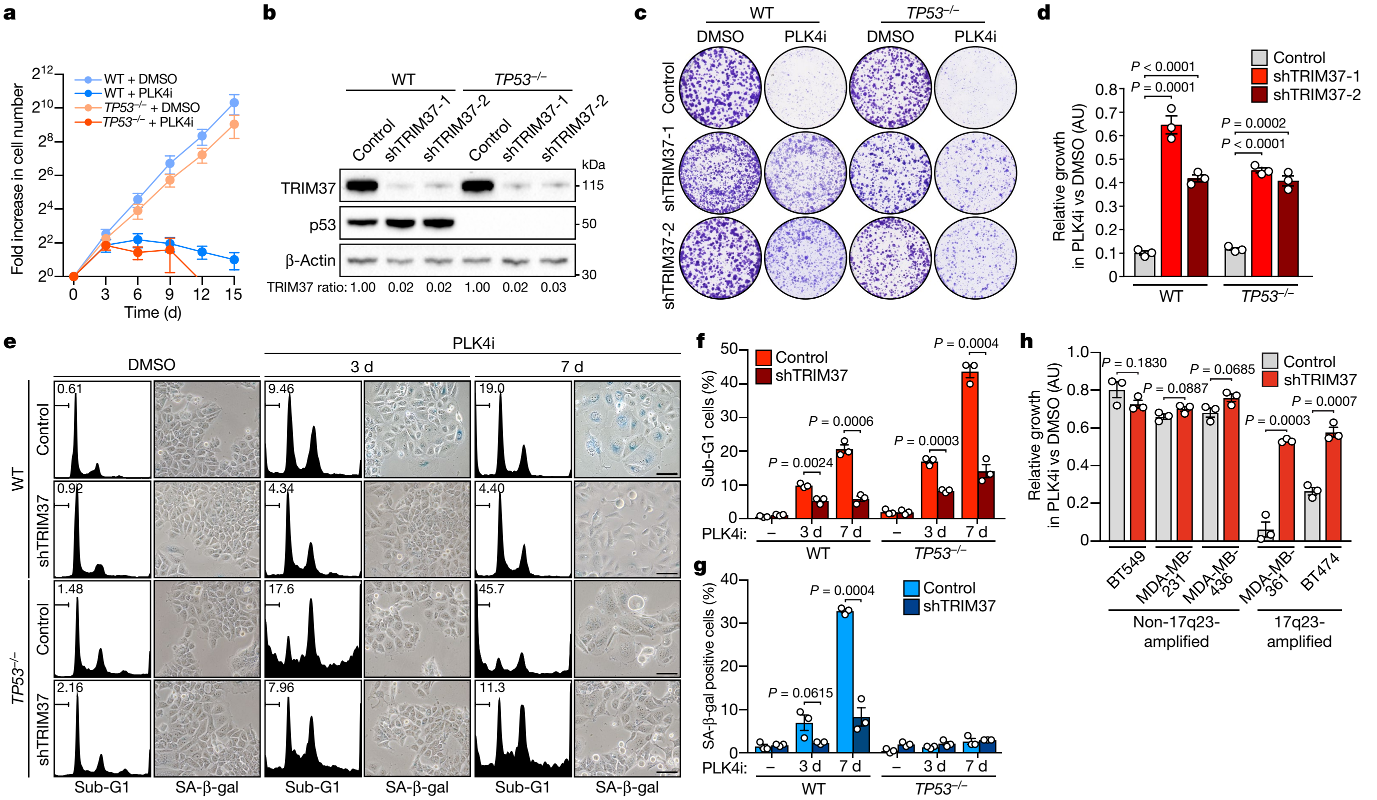


Fig. 1 | PLK4 inhibition is synthetically lethal with TRIM37 amplification. **a**, Fold increase in MCF-7 cell number after addition of centrinone (125 nM). *n* = 3 biological replicates. Mean ± s.e.m. WT, wild type. **b**, Immunoblot showing TRIM37 protein levels in wild-type and *TP53*$^{-/-}$ MCF-7 cells that stably express a control or one of two independent *TRIM37*-targeting shRNAs (shTRIM37-1 or shTRIM37-2). β-Actin, loading control. Representative data; *n* = 3 biological replicates. For gel source data, see Supplementary Fig. 1. **c**, Representative data of a 10-d clonogenic survival of indicated MCF-7 cell lines treated with DMSO (control) or centrinone (PLK4 inhibitor (PLK4i)) (125 nM). **d**, Quantification of **c**. *n* = 3 biological replicates. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **e**, MCF-7 cells treated with DMSO or centrinone (PLK4i) (125 nM) were analysed for DNA content, and stained for expression of senescence-associated β-galactosidase (SA-β-gal). Representative data of *n* = 3 biological replicates. Scale bars, 100 μm. **f**, Percentage of sub-G1 cells from **e**. *n* = 3 biological replicates. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. –, DMSO treatment. **g**, Quantification of the percentage of SA-β-gal-positive cells from **e**. *n* = 3 biological replicates, each comprising ≥200 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **h**, Quantification of clonogenic survival data for 17q23-amplified and non-17q23-amplified breast cancer cells transduced with a *TRIM37*-targeting shRNA or vector control and treated with DMSO or centrinone (PLK4i) (125 nM). *n* = 3 biological replicates. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m.

Knockout of TRIM37 leads to the accumulation of pericentriolar material (PCM) and accelerated spindle assembly in acentrosomal cells[11]. We therefore hypothesized that, conversely, high levels of TRIM37 could reduce PCM-mediated nucleation of microtubules, and thereby sensitize cells to centrosome loss. To test this, we transduced wild-type and *TP53*$^{-/-}$ MCF-7 cell lines with lentiviruses that encode control or *TRIM37*-targeting short hairpin RNAs (shRNAs), and monitored growth in the presence or absence of centrinone. In both cell lines, efficient TRIM37 depletion using two different shRNAs restored cell growth in the presence of centrinone, when compared to controls (Fig. 1b–d). Similarly, disruption of *TRIM37* with CRISPR–Cas9 conferred resistance to centrinone in two MCF-7 clones (Extended Data Fig. 1c–e). Treatment with centrinone induced senescence in MCF-7 cells, as evidenced by a time-dependent increase in cell flattening and senescence-associated β-galactosidase expression, or cell death, as marked by an accumulation of cells with a sub-G1 DNA content (Fig. 1e–g). Treatment with centrinone also inhibited the proliferation of *TP53*$^{-/-}$ MCF-7 cells, primarily by inducing cell death (Fig. 1e, f). Centrinone-induced senescence or cell death in MCF-7 cell lines was suppressed by the depletion of TRIM37 (Fig. 1e–g), which suggests that increased expression of *TRIM37* expression is synthetically lethal with PLK4 inhibition.

To test whether TRIM37 overexpression sensitizes cells to PLK4 inhibition, control (*eGFP*) or *TRIM37* transgenes were introduced into HCT116, a human colorectal carcinoma cell line that is insensitive to centrosome loss[7]. Overexpression of TRIM37 at levels comparable to those in MCF-7 cells (Extended Data Fig. 1f) inhibited clonogenic survival in centrinone-treated HCT116 cells, but only modestly affecting the growth of controls treated with dimethyl sulfoxide (DMSO) (Extended Data Fig. 1g, h). To ascertain whether the synthetically lethal effect of centrinone was specific to PLK4 inhibition, we overexpressed *TRIM37* in *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells, which exclusively express analogue-sensitive (AS) PLK4[13]. In these cells, neither doxycycline-induced *TRIM37* overexpression alone nor the inhibition of analogue-sensitive PLK4 with the bulky ATP analogue 3MB-PP1 affected cell proliferation (Extended Data Fig. 1i–l). By contrast, treatment with 3MB-PP1 resulted in cell flattening and markedly reduced colony survival in *TRIM37*-overexpressing *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells, but not in cells that overexpress a control *GST* transgene (Extended Data Fig. 1i–l). This confirms that the specific inhibition of PLK4 can kill cells with increased *TRIM37* expression.

CFI-400945 is a PLK4 inhibitor that also targets aurora B, and is in clinical trials as a therapeutic agent for patients with breast cancer[20–22]. We therefore tested the effect of CFI-400945 on the proliferation of MCF-7 cells. Treatment with centrinone, CFI-400945 or the aurora B inhibitor ZM44749 all potently inhibited clonogenic survival in MCF-7 cells (Extended Data Fig. 2a, b). However, depletion of TRIM37 restored only the proliferation of cells treated with centrinone—and not those treated with CFI-400945 or ZM447439. In cells treated with

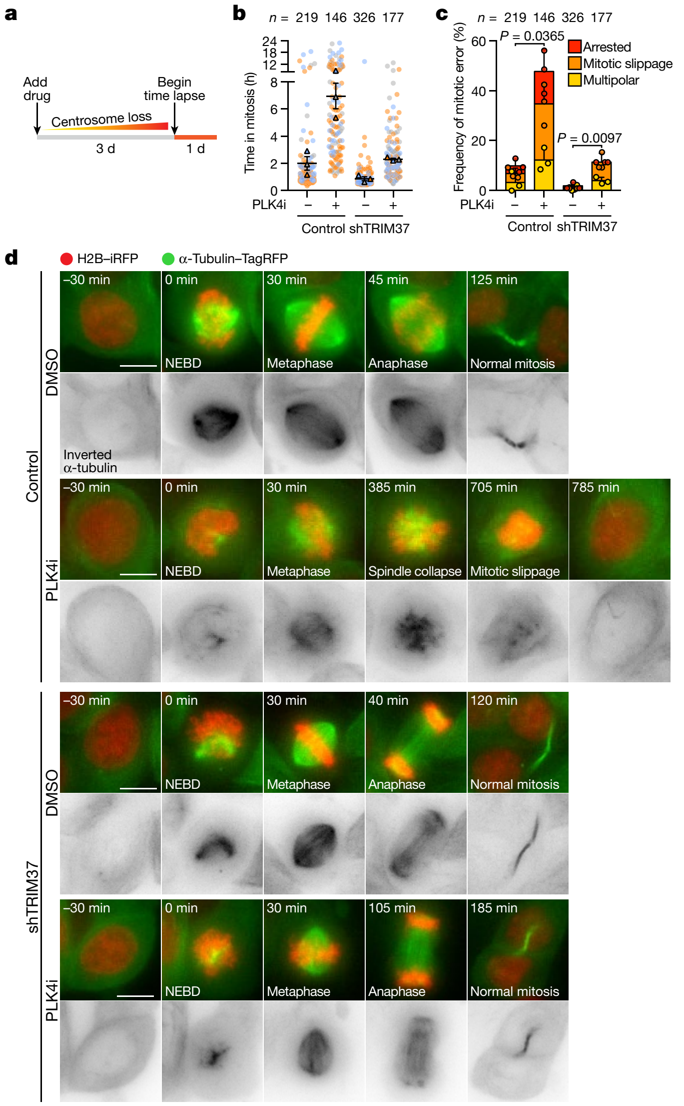


Fig. 2 | PLK4 inhibition triggers mitotic catastrophe in TRIM37-amplified cancer cells. a, Experimental schematic of time-lapse imaging of asynchronous cells. **b**, Quantification of mitotic duration in *TP53*$^{-/-}$ MCF-7 cells transduced with a control vector, compared to those expressing an shRNA targeting *TRIM37*. Cells were treated with DMSO (−) or centrinone (PLK4i) (125 nM) for 3 d before imaging. Triangles represent the mean for each biological replicate; coloured circles show individual data points from each of the replicates. Data acquired from $n = 3$ biological replicates, each with >40 cells. Mean ± s.e.m. **c**, Quantification of mitotic phenotypes from **b** in *TP53*$^{-/-}$ MCF-7 cells expressing control vector, compared to those expressing an shRNA targeting *TRIM37*. Data acquired from $n = 3$ biological replicates, each with >40 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **d**, Representative time-lapse images of mitotic progression in DMSO- or centrinone (125 nM)-treated *TP53*$^{-/-}$ MCF-7 cells expressing control vector or *TRIM37*-targeting shRNA. Representative data; $n = 3$ biological replicates. Cells are labelled with H2B–iRFP and TagRFP–α-tubulin. NEBD, nuclear envelope break down. Scale bars, 5 μm.

CFI-400945 or ZM447439, analysis of DNA content revealed an accumulation of polyploid cells, which is a readout indicative of inhibition of aurora B kinase (Extended Data Fig. 2c). By contrast, treatment with centrinone did not increase the fraction of polyploid cells, consistent with the known selectivity of this compound for inhibiting PLK4 and not aurora B[7]. This shows that inhibitor selectivity towards PLK4—and not other kinases—is crucial for the synthetic lethal effect in cells that overexpress *TRIM37*.

To determine whether the cell killing induced by PLK4 inhibitors was common to 17q23-amplified breast cancer cell lines, we tested the effect of centrinone on the viability of BT474 and MDA-MB-361 (two additional 17q23-amplified breast cancer cell lines that overexpress *TRIM37*[16,23]), and compared their responses with those of a control panel of non-17q23-amplified breast cancer cell lines (BT549, MDA-MB-231 and MDA-MB-436) with normal TRIM37 expression (Extended Data Fig. 3a). As expected, PLK4 inhibition only minimally affected clonogenic survival across the control cell line panel, and TRIM37 depletion conferred no added resistance (Fig. 1h, Extended Data Fig. 3b). By contrast, both of the additional 17q23-amplified cell lines were hypersensitive to the PLK4 inhibitor; treatment with centrinone induced growth arrest, morphological aberrations and cell death (Fig. 1h, Extended Data Fig. 3b–d). These effects were also suppressed by stable knockdown of TRIM37, which confirms that the synthetically lethal effect of PLK4 inhibitor treatment depended on *TRIM37* overexpression in multiple 17q23-amplified cell lines.

To test whether TRIM37 overexpression was predictive for sensitivity to PLK4 inhibitors in patient-derived organoid models of breast cancer, we examined the centrinone sensitivity of 3D organoid cultures derived from patients with breast cancer with a high or low level of expression of *TRIM37* mRNA (Extended Data Fig. 3e). Expression of *TRIM37* mRNA was only partly predictive of TRIM37 protein levels (Extended Data Fig. 3g). Nevertheless, of four established cultures, two patient-derived organoids with high levels of TRIM37 protein—and one with an intermediate level of TRIM37 protein—were sensitive to nanomolar doses of centrinone. By contrast, a patient-derived organoid with low levels of TRIM37 protein remained insensitive to treatment with centrinone at concentrations below 1 μM (Extended Data Fig. 3f, g). The centrinone-sensitivity profiles of these patient-derived organoids resembled the 3D-culture responses of 17q23-amplified (MCF-7 and BT-474) and non-amplified (MDA-MB-231 and BT-549) breast tumour cell lines to centrinone treatment (Extended Data Fig. 3h). Taken together, our cell line and patient-derived-organoid experiments underscore the utility of PLK4-specific inhibitors in the killing of TRIM37-amplified breast cancer cells.

To investigate how PLK4 inhibition triggers growth defects in MCF-7 cells, we performed time-lapse microscopy to track the fates of control (DMSO)- and centrinone-treated *TP53*$^{-/-}$ MCF-7 cells (Fig. 2a, Supplementary Videos 1–4). Whereas control-treated cells progressed through mitosis normally, 47% of centrinone-treated cells formed short bipolar spindles that collapsed and remained arrested in mitosis or slipped out of mitosis without undergoing anaphase (Fig. 2b–d). Importantly, TRIM37 depletion rescued robust bipolar spindle formation in centrinone-treated MCF-7 cells, and almost completely reversed the effects of centrinone in prolonging mitosis and inducing cell division errors in MCF-7 cells (Fig. 2b–d). Thus, increased expression of *TRIM37* antagonizes spindle assembly in the absence of centrosomes, resulting in mitotic catastrophe.

To understand how TRIM37 inhibits spindle assembly, we identified proximity interaction partners by expressing mTurbo-tagged TRIM37 and performing proximity-dependent biotin labelling in RPE-1 cells[24,25]. After background subtraction, we identified 184 TRIM37 proximity-interaction partners, including 7 known interactors[26] (Extended Data Fig. 4a, c, Supplementary Data 1). Gene ontology analysis showed notable enrichment of centrosome proteins within these interactors (Extended Data Fig. 4b). This was corroborated by the localization of a pool of endogenous TRIM37 in close proximity to the centrosome, and the enrichment of biotinylated proteins at the centrosomes of RPE-1 cells expressing mTurbo–TRIM37 (Extended Data Fig. 4d, e). Among the most enriched proximity interactors of TRIM37 was CEP192, a core PCM component that accumulates in non-centrosomal foci in TRIM37-knockout cells[11]. The interaction between TRIM37 and CEP192 was confirmed by co-immunoprecipitation (Extended Data Fig. 4f). Neither histone H2A nor the peroxisome protein PEX5—two previously

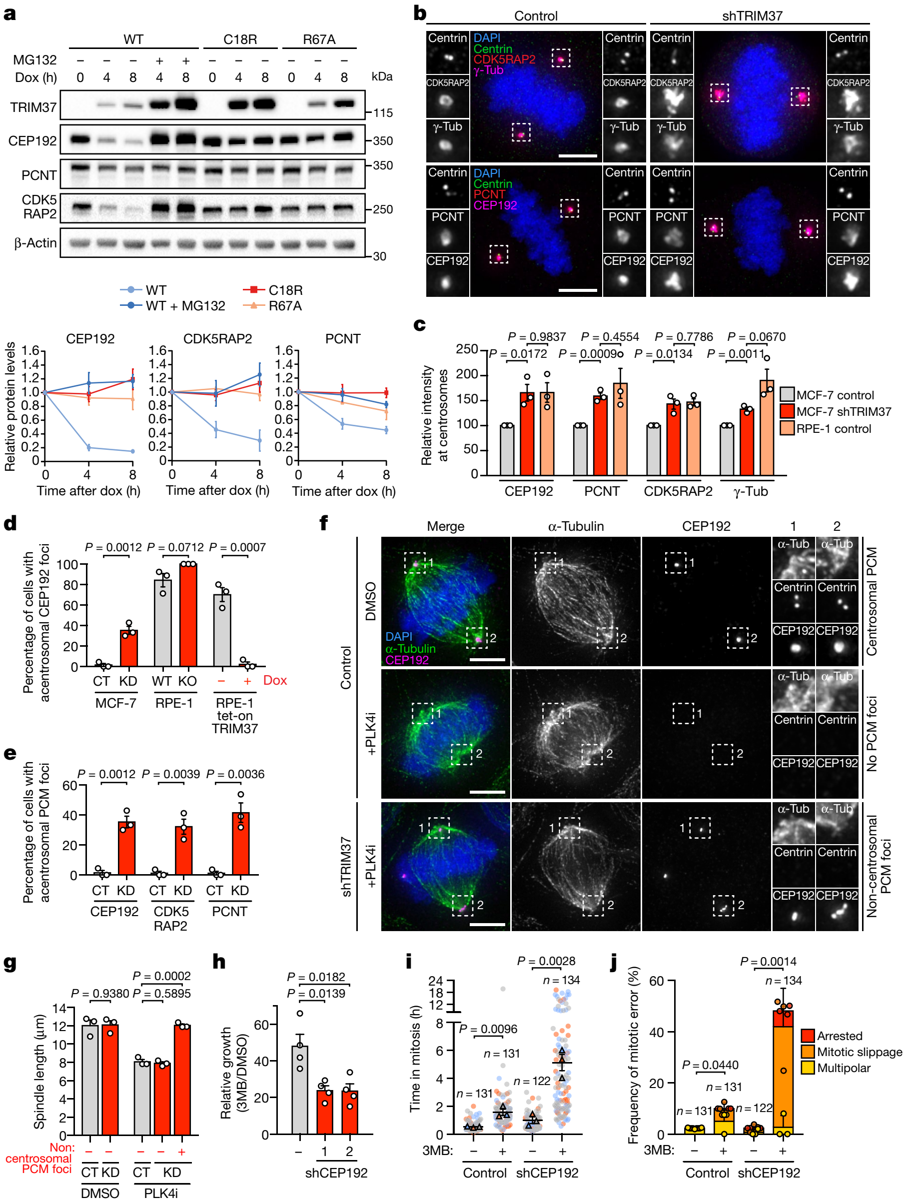


Fig. 3 | PCM sequestration by TRIM37 drives mitotic catastrophe in acentrosomal cells. a, Top, immunoblot showing PCM levels after overexpression of wild-type TRIM37 (WT), the RING-domain mutant TRIM37(C18R) or the ubiquitin-transfer-defective mutant TRIM37(R67A) in RPE-1 tet-on TRIM37 cells. β-Actin, loading control. For gel source data, see Supplementary Fig. 1. Bottom, normalized PCM levels relative to 0 h, representative of $n = 3$ biological replicates. Mean ± s.e.m. Dox, doxycycline. **b**, Centrosomal PCM levels in mitotic MCF-7 cells transduced with control vector or *TRIM37*-targeting shRNA. Representative images, $n = 3$ biological replicates. Scale bars, 5 μm. γ-Tub, γ-tubulin. **c**, Quantification of centrosomal PCM signal in mitotic cells. $n = 3$ biological replicates. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **d**, Quantification of mitotic CEP192 foci in centrinone-treated cells that lack centrosomes. $n = 3$, biological replicates, each comprising >30 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. CT, control vector; KD, knockdown with *TRIM37* shRNA; KO, *TRIM37* knockout. **e**, Quantification of mitotic PCM foci in centrinone-treated cells that lack centrosomes. $n = 3$ biological replicates, each comprising >30 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **f**, Representative images for **e**. Scale bars, 5 μm. **g**, Quantification of mitotic spindle length in MCF-7 cells expressing control vector or *TRIM37*-targeting shRNA. $n = 3$ biological replicates, each comprising >10 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **h**, Growth of 3MB-PP1 (3MB)-treated *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells expressing control vector or one of two *CEP192*-targeting shRNAs, relative to treatment with DMSO. $n = 4$ biological replicates. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **i**, Mitotic duration of the cells described in **h**, expressing H2B–eGFP and TagRFP–tubulin. Cells were grown in DMSO or 3MB-PP1 for 3 d before imaging. Triangles, mean for each biological replicate; coloured circles, individual data points from each replicate. $n = 3$ biological replicates, each comprising >30 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **j**, Frequency of mitotic errors quantified in the samples as described in **i**. $n = 3$ biological replicates, each comprising >30 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m.

reported substrates of TRIM37[23,27]—were among the labelled interactors, which confirms that centrosome proteins are primary TRIM37 proximity-interaction partners.

To test whether TRIM37 regulates the abundance of PCM proteins, we monitored the effect of altered TRIM37 expression on the cell-wide levels of three PCM scaffolding proteins: CEP192, PCNT and CDK5RAP2. Acute overexpression of *TRIM37* in RPE-1 cells markedly reduced the abundance of all three PCM proteins (Fig. 3a). Proteasome blockade with MG132 prevented the reduction in PCM protein abundance, which suggests that TRIM37 directs the degradation of these proteins via the ubiquitin-proteasome pathway. Consistent with this, the E3 ligase activity of TRIM37 was critical for PCM protein degradation, as both catalytically inactive (C18R)[23,27] and predicted ubiquitin-binding- and transfer-defective (R67A)[28] RING domain mutants of TRIM37 did not reduce levels of the PCM proteins (Fig. 3a). Having established that TRIM37 directs PCM protein proteolysis, we next investigated whether 17q23-amplification status correlated with reduced PCM levels. We found that the 17q23-amplified cell lines with TRIM37 overexpression have lower cell-wide levels of CEP192, PCNT and CDK5RAP2 compared to their non-17q23-amplified counterparts (Extended Data Fig. 4g). The levels of CEP192, PCNT and CDK5RAP2 were also reduced at mitotic centrosomes in MCF-7 cells, but restored following TRIM37 depletion to levels comparable to those seen in RPE-1 cells (Fig. 3b, c). Microtubule regrowth assays showed that mitotic centrosomes in TRIM37-depleted MCF-7 cells nucleated nearly twice the amount of α-tubulin compared to the control cells (Extended Data Fig. 5a, b). Similarly, the levels of EB1, a plus-end tracking marker of growing microtubules, were increased by more than threefold at the centrosomes of TRIM37-depleted cells (Extended Data Fig. 5a, c).

Fixed-cell analysis revealed that acentrosomal RPE-1, DLD-1 and MDA-MB-436 cells that express low levels of TRIM37 formed PCM aggregates in more than 80% of mitotic cells. However, these PCM foci were absent from MCF-7 cells and from TRIM37-overexpressing RPE-1 cells (Fig. 3d–f, Extended Data Fig. 5d, e). We therefore asked whether the role of TRIM37 in controlling PCM abundance could modulate the assembly of non-centrosomal PCM foci. Depleting TRIM37 enabled the formation of PCM foci in acentrosomal MCF-7 cells (Fig. 3d–f) and increased the penetrance and size of these structures in acentrosomal RPE-1 cells (Fig. 3d, Extended Data Fig. 5d, f). Centrosome loss also reduced the length of the mitotic spindle in MCF-7 cells (Fig. 3g, Extended Data Fig. 5g). However, TRIM37 depletion enabled acentrosomal MCF-7 cells with PCM foci to generate spindle lengths that matched those of untreated MCF-7 cells. Thus, *TRIM37* overexpression suppresses the formation of non-centrosomal PCM foci, which leads to defects in spindle assembly.

To define the spatial and temporal assembly properties of non-centrosomal PCM foci, we generated DLD-1 cells that express endogenously tagged CEP192–mNeonGreen to mark PCM. In DMSO-treated control DLD-1 cells, CEP192–mNeonGreen localized to the centrosomes in interphase and increased in intensity by about threefold during mitosis (Extended Data Fig. 6a, Supplementary Video 5). By contrast, CEP192–mNeonGreen was diffusely localized in acentrosomal DLD-1 cells throughout interphase but assembled into multiple non-centrosomal PCM foci during early prometaphase (Extended Data Fig. 6b, Supplementary Video 6). These PCM foci coalesced into spindle poles at metaphase, and subsequently disassemble upon mitotic exit.

Non-centrosomal PCM foci often resided at the centre of microtubule asters (Extended Data Fig. 5d, e, g). To examine whether these structures could promote microtubule nucleation, we performed live-cell confocal imaging on acentrosomal DLD-1 cells that express endogenously tagged CEP192–mNeonGreen and EB1–TagRFP. In DMSO-treated control cells, EB1–TagRFP tracked the growing ends of microtubules nucleated by the centrosomes (Extended Data Fig. 6c, Supplementary Video 7). In acentrosomal DLD-1 cells, non-centrosomal PCM foci nucleated microtubules and were incorporated into the mitotic spindle (Extended Data Fig. 6d, Supplementary Video 8). Importantly, microtubule nucleation by non-centrosomal PCM foci preceded their incorporation into the spindle (Extended Data Fig. 6e, Supplementary Video 9). Similar results were obtained with acentrosomal RPE-1 cells that co-express EB3–mNeonGreen to track the plus-end tip of microtubules and γ-tubulin–TagRFP to mark acentrosomal PCM foci (Extended Data Fig. 7, Supplementary Videos 10–14). We conclude that, in cells that lack centrosomes, non-centrosomal PCM foci form specifically in mitosis and are required for efficient microtubule nucleation and robust bipolar spindle assembly.

To test whether PCM depletion could explain the synthetically lethal effect of centrosome loss in cells with *TRIM37* overexpression, we depleted the TRIM37 target CEP192 in $PLK4^{AS}$ $TP53^{-/-}$ RPE-1

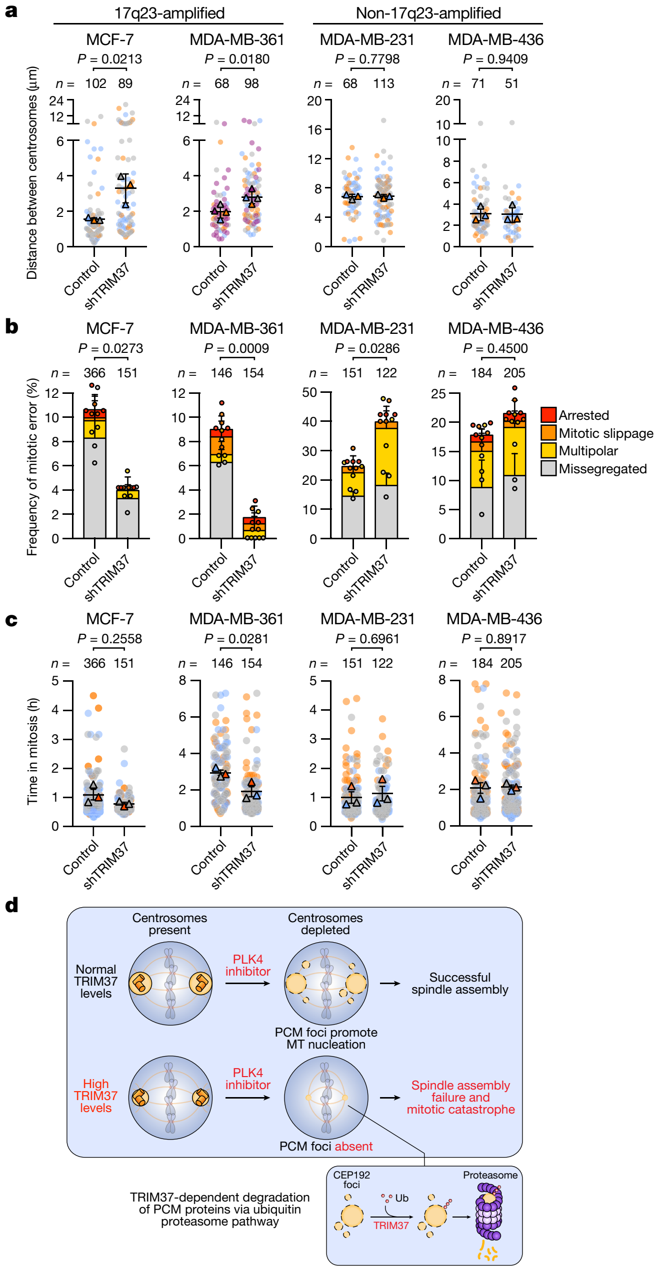


Fig. 4 | TRIM37 overexpression delays centrosome separation in late G2 phase, and promotes mitotic errors. a, Quantification of the distance between the two centrosomes at NEBD in MCF-7, MDA-MB-361, MDA-MB-231 and MDA-MB-436 cells expressing a control vector, compared to those expressing an shRNA targeting *TRIM37*. Data acquired from $n \geq 3$ biological replicates, each with 8–45 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **b**, Quantification of mitotic phenotypes in MCF-7, MDA-MB-361, MDA-MB-231 and MDA-MB-436 cells expressing a control vector, compared to those expressing an shRNA targeting *TRIM37*. Data acquired from $n = 3$ biological replicates, each with >40 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **c**, Quantification of mitotic duration in MCF-7, MDA-MB-361, MDA-MB-231 and MDA-MB-436 cells expressing a control vector, compared to those expressing an shRNA targeting *TRIM37*. Triangles represent the mean for each biological replicate; coloured circles show individual data points from each of the replicates. Data acquired from $n = 3$ biological replicates, each with >40 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **d**, A model illustrating the synthetic lethal effect of PLK4 inhibition with TRIM37 overexpression in 17q23-amplified breast cancer cells. MT, microtubule; Ub, ubiquitin.

(Extended Data Fig. 8a–c). As predicted, depletion of CEP192 sensitized these cells to centrosome loss (Fig. 3h). Treatment with 3MB-PP1 in CEP192-depleted RPE-1 cells also recapitulated the mitotic phenotypes we observed in acentrosomal MCF-7 cells, including a reduced frequency of non-centrosomal CEP192 foci, prolonged mitotic duration and a marked increase in mitotic errors (Fig. 3i, j, Extended Data Fig. 8d, e, Supplementary Videos 15–18). We conclude that TRIM37-dependent PCM depletion in mitosis leads to delayed and inefficient microtubule nucleation; these represent defects that can be exploited to induce mitotic catastrophe upon pharmacological depletion of centrosomes.

17q23-amplicon-positive breast cancer typically comprises highly proliferative, ER-positive luminal B-type tumours, characterized by high levels of genomic instability[4,5]. However, the mechanisms and driver genes that are responsible for genomic instability in 17q23-amplified tumours remain undefined. The high burden of genomic instability in MCF-7 cells drives their rapid genetic diversification in culture[29], which led us to consider the contribution of *TRIM37* overexpression to this process. Previous work has shown that delayed centrosome separation increases the rates of kinetochore mis-attachment and of mitotic errors in cancer cells[30–32]. We therefore investigated whether increased *TRIM37* expression during mitosis modulates the timing of centrosome maturation and separation. In contrast to RPE-1 cells, in which TRIM37 was transcriptionally downregulated in G2 phase and mitosis (Extended Data Fig. 9a–e), expression of TRIM37 protein persisted at high levels throughout the cell cycle in MCF-7 cells (Extended Data Fig. 9d). TRIM37 depletion in MCF-7 cells accelerated centrosome maturation by 20 min in late G2 phase (Extended Data Fig. 10a, d, Supplementary Videos 19, 20), and increased centrosome separation at mitotic entry (Extended Data Fig. 10b). Conversely, TRIM37 overexpression in RPE-1 cells delayed centrosome maturation in late G2 phase by 17 min (Extended Data Fig. 10c, e, Supplementary Videos 21, 22). Collectively, these data show that TRIM37-driven suppression of PCM assembly delays microtubule nucleation and centrosome separation in late G2 phase.

To test whether delays in centrosome separation at mitotic entry dependent on TRIM37 overexpression could cause mitotic errors, we compared the effect of the level of TRIM37 expression on the timing of centrosome separation, mitotic duration and the frequency of mitotic errors in 17q23-amplified (MCF-7 and MDA-MB-361) versus non-amplified (MDA-MB-231 and MDA-MB-436) breast cancer cell lines. TRIM37 depletion increased the distance between the centrosomes at early prophase and reduced mitotic duration in the 17q23-amplified cancer cell lines (Fig. 4a, c). By contrast, TRIM37 knockdown did not alter centrosome separation timing or mitotic duration in non-17q23-amplified breast cancer cells (Fig. 4a, c). Importantly, TRIM37 depletion also reduced the frequency of mitotic errors in 17q23-amplified breast cancer cells, but trended towards increasing the rate of cell division errors in non-17q23-amplified cancer cells (Fig. 4b). These data show that overexpression of TRIM37 delays the timing of centrosome separation at mitotic entry in 17q23-amplified breast cancer cells, and suggest that this delay in centrosome separation promotes genetic instability by increasing the frequency of mitotic errors.

We propose that TRIM37 usually acts to inhibit the assembly of non-centrosome-associated PCM into structures that would otherwise compromise mitotic fidelity. Consequently, the centrosomal defects that accompany *TRIM37* amplification may fuel the stochastic mitotic errors that contribute to the high burden of genomic instability in 17q23-amplified breast tumours and could drive tumour evolution. We also propose centrosome depletion as a therapeutic strategy to kill cancers that overexpress *TRIM37*. High levels of TRIM37 reduce the availability of PCM proteins, thereby impeding the formation of non-centrosomal PCM foci—assemblies that we propose are required for mitosis in the absence of centrosomes (Fig. 4d). Our work therefore indicates that the inhibition of PLK4, or other regulators of centrosome duplication or assembly, represents a promising strategy to selectively target breast cancers or other tumours[33] driven by 17q23 amplification.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2690-1.

1. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345 (2013).
2. Kalimutho, M. et al. Patterns of genomic instability in breast cancer. *Trends Pharmacol. Sci*. **40**, 198–211 (2019).
3. Lord, C. J. & Ashworth, A. PARP inhibitors: synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017).
4. Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
5. Dawson, S.-J., Rueda, O. M., Aparicio, S. & Caldas, C. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J*. **32**, 617–628 (2013).
6. Ali, H. R. et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. *Genome Biol*. **15**, 431 (2014).
7. Wong, Y. L. et al. Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science* **348**, 1155–1160 (2015).
8. Moyer, T. C., Clutario, K. M., Lambrus, B. G., Daggubati, V. & Holland, A. J. Binding of STIL to Plk4 activates kinase activity to promote centriole assembly. *J. Cell Biol*. **209**, 863–878 (2015).
9. Lambrus, B. G. et al. p53 protects against genome instability following centriole duplication failure. *J. Cell Biol*. **210**, 63–77 (2015).
10. Cuella-Martin, R. et al. 53BP1 integrates DNA Repair and p53-dependent cell fate decisions via distinct mechanisms. *Mol. Cell* **64**, 51–64 (2016).
11. Meitinger, F. et al. 53BP1 and USP28 mediate p53 activation and G1 arrest after centrosome loss or extended mitotic duration. *J. Cell Biol*. **214**, 155–166 (2016).
12. Fong, C. S. et al. 53BP1 and USP28 mediate p53-dependent cell cycle arrest in response to centrosome loss and prolonged mitosis. *eLife* **5**, e16270 (2016).
13. Lambrus, B. G. et al. A USP28–53BP1–p53–p21 signaling axis arrests growth after centrosome loss or prolonged mitosis. *J. Cell Biol*. **214**, 143–153 (2016).
14. Ciriello, G. et al. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* **163**, 506–519 (2015).
15. Pereira, B. et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat. Commun*. **7**, 11479 (2016).
16. Monni, O. et al. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl Acad. Sci. USA* **98**, 5711–5716 (2001).
17. Pärssinen, J., Kuukasjärvi, T., Karhu, R. & Kallioniemi, A. High-level amplification at 17q23 leads to coordinated overexpression of multiple adjacent genes in breast cancer. *Br. J. Cancer* **96**, 1258–1264 (2007).
18. Sinclair, C. S., Rowley, M., Naderi, A. & Couch, F. J. The 17q23 amplicon and breast cancer. *Breast Cancer Res. Treat*. **78**, 313–322 (2003).
19. Balestra, F. R., Strnad, P., Flückiger, I. & Gönczy, P. Discovering regulators of centriole biogenesis through siRNA-based functional genomics in human cells. *Dev. Cell* **25**, 555–571 (2013).
20. Oegema, K., Davis, R. L., Lara-Gonzalez, P., Desai, A. & Shiau, A. K. CFI-400945 is not a selective cellular PLK4 inhibitor. *Proc. Natl Acad. Sci. USA* **115**, E10808 (2018).
21. Mason, J. M. et al. Functional characterization of CFI-400945, a Polo-like kinase 4 inhibitor, as a potential anticancer agent. *Cancer Cell* **26**, 163–176 (2014).
22. Kawakami, M. et al. Polo-like kinase 4 inhibition produces polyploidy and apoptotic death of lung cancers. *Proc. Natl Acad. Sci. USA* **115**, 1913–1918 (2018).
23. Bhatnagar, S. et al. TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. *Nature* **516**, 116–120 (2014).
24. Branon, T. C. et al. Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol*. **36**, 880–887 (2018).
25. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol*. **196**, 801–810 (2012).
26. Oughtred, R. et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res*. **47**, D529–D541 (2019).
27. Wang, W., Xia, Z.-J., Farré, J.-C. & Subramani, S. TRIM37, a novel E3 ligase for PEX5-mediated peroxisomal matrix protein import. *J. Cell Biol*. **216**, 2843–2858 (2017).
28. Densham, R. M. et al. Human BRCA1–BARD1 ubiquitin ligase activity counteracts chromatin barriers to DNA resection. *Nat. Struct. Mol. Biol*. **23**, 647–655 (2016).
29. Ben-David, U. et al. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560**, 325–330 (2018).
30. Silkworth, W. T., Nardi, I. K., Paul, R., Mogilner, A. & Cimini, D. Timing of centrosome separation is important for accurate chromosome segregation. *Mol. Biol. Cell* **23**, 401–411 (2012).
31. Kaseda, K., McAinsh, A. D. & Cross, R. A. Dual pathway spindle assembly increases both the speed and the fidelity of mitosis. *Biol. Open* **1**, 12–18 (2012).
32. Zhang, Y. et al. USP44 regulates centrosome positioning to prevent aneuploidy and suppress tumorigenesis. *J. Clin. Invest*. **122**, 4362–4374 (2012).
33. Meitinger, F. et al. TRIM37 controls cancer-specific vulnerability to PLK4 inhibition. *Nature* https://doi.org/10.1038/s41586-020-2710-1 (2020).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Cell lines and culture conditions

MCF-7, MDA-MB-231, MDA-MB-436, DLD-1 and HCT-116 cells were grown in DMEM medium (Corning Cellgro) containing 10% fetal bovine serum (Sigma), 100 U/ml penicillin, 100 U/ml streptomycin and 2 mM L-glutamine. hTERT RPE-1 cells were grown in DMEM:F12 medium (Corning Cellgro) containing 10% fetal bovine serum (Sigma), 0.348% sodium bicarbonate, 100 U/ml penicillin, 100 U/ml streptomycin and 2 mM L-glutamine. MDA-MB-361 cells were grown in DMEM medium (ThermoFisher Scientific) containing 20% fetal bovine serum (Sigma), 100 U/ml penicillin, 100 U/ml streptomycin and 2 mM L-glutamine. BT474 and BT549 cells were grown in RPMI 1640 medium (ThermoFisher Scientific) containing 10% fetal bovine serum (Sigma), 100 U/ml penicillin, 100 U/ml streptomycin and 2 mM L-glutamine and 10 µg/ml bovine insulin (Sigma). All cell lines were maintained at 37 °C in a 5% CO_2 atmosphere with 21% oxygen and routinely checked for mycoplasma contamination.

Centrinone sensitivity in 3D tumour cell line and patient-derived organoid (PDO) cultures

Human breast tumour samples were obtained from adult female patients after informed consent as part of a non-interventional clinical trial (BTBC study REC no.: 13/LO/1248, IRAS ID 131133; principal investigator: A.N.J.T., study title: 'Analysis of functional immune cell stroma and malignant cell interactions in breast cancer in order to discover and develop diagnostics and therapies in breast cancer subtypes'). This study had local research ethics committee approval and was conducted adhering to the principles of the Declaration of Helsinki. Specimens were collected from surgery and transported immediately. A clinician histopathologist or pathology-trained technician identified and collected tumour material into basal culture medium. Tumour samples were coarsely minced with scalpels and then dissociated using a Gentle MACS dissociator (Miltenyi). The resulting cell suspension was mechanically disrupted, filtered and centrifuged. Resulting cell pellets were then plated into 3D cultures at approximately 1×10^3 to 2×10^3 cells per µl in Ocello PDX medium (OcellO B.V) and hydrogel as previously described[34–36]. All cultures were maintained in humidified incubators at 37 °C, 5% CO_2. For centrinone sensitivity analysis, PDOs (between passage 10 and 25) and tumour cell lines (short-tandem-repeat-typed every five passages to confirm identity) were dissociated to single-cell populations using TrypLE (Life Technologies). Cell suspensions were then dispensed into 384-well plates in Ocello PDX medium and hydrogel. Twenty-four h after seeding, cultures were treated with centrinone (diluted in 0.1% v/v DMSO) and then continuously cultured for a total of 13 d, with drug-containing medium being replenished every 4 d. After this point, cell viability was estimated using Cell Titer Glo 3D (CTG, Promega) as per the manufacturer's guidelines. CTG luminescence was measured using a Victor X5 plate reader (Perkin Elmer). Data are presented as per cent survival normalized to cells exposed to 0.1% v/v DMSO alone.

Gene targeting and stable cell lines

To generate CRISPR–Cas9-mediated knockout lines, gene-specific sgRNAs (*TP53Δ*, 5′-gtgcagctgtgggttgattc-3′; *TP53BP1Δ*, 5′-gaacgaggagacggtaatagt-3′; *USP28Δ*, 5′-tgccattgctttgagtctac-3′) were cloned into a modified pX330 vector (no. 42230; Addgene) containing a puromycin-resistance cassette. Cells were transiently transfected (Fugene HD, Promega) with the pX330 plasmids and positive selection of transfected cells was performed 2 d after transfection with 2.0 µg/ml puromycin. Monoclonal cell lines were isolated by limiting dilution. The presence of gene-disrupting insertions and deletions (indels) in edited cell lines was confirmed by Sanger sequencing, and the ablation of protein production was assessed by immunoblotting.

To generate EB1–TagRFP-, H2B–iRFP- and TagRFP–tubulin- or eGFP–tubulin-labelled cell lines, ORFs were cloned into FUGW lentiviral vectors. Fluorescent populations of cells were generated by lentivirus-mediated transduction. MCF-7 cells were transduced with H2B–iRFP and TagRFP–tubulin. RPE-1 cells were transduced with H2B–iRFP and eGFP–tubulin. DLD-1 cells expressing CEP192–mNeonGreen cells were transduced with H2B–iRFP and EB1–TagRFP. MDA-MB-361, MDA-MB-231 and MDA-MD-436 cells were transduced with H2B–iRFP. Polyclonal populations of cells expressing the desired fluorescent markers were used directly or isolated using FACS.

To generate EB3–mNeonGreen- and γ-tubulin–TagRFP-labelled cell lines, the two open reading frames (ORFs) separated by a T2A sequence were cloned into a CMV-puro lentiviral vector. RPE-1 cells were transduced with EB3–mNeonGreen and γ-tubulin–TagRFP dual-expressing lentivirus, and polyclonal populations of cells expressing both markers were selected using puromycin.

To generate *TRIM37*-overexpressing cell lines, the *TRIM37* ORF was cloned into a constitutive or tet-inducible lentiviral vector. The C18R and R67A mutations were introduced using PCR-directed mutagenesis and verified by Sanger sequencing. Cells were transduced and stable polyclonal populations of cells selected and maintained in the presence of 1.0 µg/ml puromycin.

To create the DLD-1 cell line expressing CEP192–mNeonGreen, an sgRNA targeting the *CEP192* translational stop codon (5′- cgactaattggtgaagctct-3′) was cloned into a pX459 vector (no. 62988; Addgene). To generate the CEP192 repair vector, we cloned a 2× mNeonGreen tag followed by a T2A-neomycin and a translational stop codon into a modified pUC vector. The 475-bp 5′ and 462-bp 3′ homology arms were PCR-amplified from genomic DLD-1 DNA and cloned on either side of the central 2× mNeonGreen-T2A-neomycin cassette. DLD-1 cells were transiently transfected (X-tremeGENE HP, Roche) with the pX459 plasmid and repair vector. Selection of transfected cells was performed 5 d after transfection with 400 µg/ml G418.

RNA interference

shRNAs targeting *TRIM37* (*TRIM37*-1, 5′-tcgagaatatgatgctgtg-3′; *TRIM37*-2, 5′-aggactttgctggaggtta-3′) were cloned into the pGIPz (Thermofisher Scientific) vector. shRNAs targeting *CEP192* (*CEP192*-1, 5′-cctgttacataaaccagagat-3′; *CEP192*-2, 5′-gaggcatcagttaatactgat-3′) were cloned into pLKO.1. Stable shRNA-mediated knockdown cell lines were generated by lentivirus-mediated transduction. Polyclonal populations of cells were subsequently selected and maintained in the presence of puromycin (1.0 µg/ml). Knockdown efficiency was assessed by immunoblotting.

Lentiviral production and transduction

Lentiviral expression vectors were cotransfected into 293FT cells with the lentiviral packaging plasmids psPAX2 and pMD2.G (Addgene no. 12260 and no. 12259). In brief, 3×10^6 293FT cells were seeded into a poly-L-lysine-coated 10-cm culture dish the day before transfection. For each 10-cm dish, the following DNA was diluted in 0.6 ml of OptiMEM (Thermo Fisher Scientific): Four and half µg of lentiviral vector, 6 µg of psPAX2 and 1.5 µg of pMD2.G. Separately, 72 µl of 1 µg/µl 25 kDa polyethylenimine (PEI; Sigma) was diluted into 1.2 ml of OptiMEM, briefly vortexed and incubated at room temperature for 5 min. After incubation, the DNA and PEI mixtures were combined, briefly vortexed and incubated at room temperature for 20 min. During this incubation, the culture medium was replaced with 17 ml of pre-warmed DMEM + 1% FBS. The transfection mixture was then added drop-wise to the 10-cm dish. Viral particles were collected 48 h after the medium change and filtered through a 0.45-µm PVDF syringe filter. The filtered supernatant was either concentrated in 100-kDa Amicon Ultra Centrifugal Filter Units (Millipore) or used directly to infect cells. Aliquots were snap-frozen and stored at −80 °C. For transduction, lentiviral particles were diluted in complete growth medium supplemented with 10 μg/ml polybrene (Sigma) and added to cells.

Chemical inhibitors

3MB-PP1 (Millipore) was dissolved in DMSO and used at a final concentration of 10 μM, and centrinone (a gift from K. Oegema) was dissolved in DMSO and used at a final concentration of 125 nM, unless otherwise indicated. CFI-400945 (Cayman Chemicals) was dissolved in DMSO and used at a final concentration of 50 or 500 nM. ZM447439 (Cayman Chemicals) was dissolved in DMSO and used at a final concentration of 2 μM. MG132 (Sigma) was dissolved in DMSO and used at a final concentration of 10 μM. CHX (VWR International) was dissolved in DMSO and used at a final concentration of 100 μg/ml. RO-3306 (Sigma) was dissolved in DMSO and used at a final concentration of 9 μM.

TRIM37 RNA abundance in PDOs

RNA from PDO cell pellets was extracted using the RNeasy kit (Qiagen) according to the manufacturer's instructions. Quality and quantity of RNA were assessed using a Qubit and Bioanalyzer (Agilent). NEBNEXT Ultra II Directional RNA and polyA RNA selection kits (Illumina) were used to generate paired-end sequencing libraries that were sequenced on an Illumina NovaSeq 6000 S2 platform. Paired-end reads were aligned to the human reference genome GRCh38 using STAR v.2.5.1b[37] using quantMode GeneCounts and twopassMode basic alignment settings. Feature quantification was performed using GENCODE (v.22) GTF file. Post alignment quality control was performed using RseQC (v.2.6.3)[38]. Data was normalized using the TMM method (trimmed mean of *M*-values) of edgeR[39] and *TRIM37* mRNA expression values converted into *Z* scores adjusted to the median of all mRNA species in the sample.

BioID sample preparation, mass spectrometry and data analysis

To generate cell lines for BioID, puro-sensitive RPE-1 cells were transduced with lentivirus containing tet-inducible miniTurbo control, or miniTurbo-TRIM37 constructs. Forty-eight h after transduction, cells were selected in 2.5 μg/ml puromycin for 2 d. Cells were then expanded into 7 × 15-cm dishes. One day before biotin labelling, 1 μg/ml doxycycline was added to induce expression of miniTurbo constructs. Twenty-four h after the start of induction, with cells at about 60% confluency, 10 μM dimethylenastron (Sigma) was added to cell culture medium to block cells in mitosis. After 2 h of mitotic block, medium was supplemented with 250 uM ᴅ-biotin (P212121; prepared as 250 mM stock in DMSO) to initiate labelling of proximity interactors. After 4 h of biotin labelling, mitotic samples were collected by mitotic shake-off, and the remaining interphase cells were collected by scraping. All samples were transferred to 15-ml conical flasks and rinsed 4 times with PBS to remove excess biotin. Cell pellets were lysed in approximately 1.5 ml lysis buffer (all buffer recipes have previously been published[40]) by gentle pipetting followed by sonication. Lysates were clarified by centrifugation at 16,000*g* for 10 min at 4 °C. To enrich for biotinylated material, 60 μl of streptavidin agarose bead resin (Pierce) was washed with lysis buffer, then incubated with clarified lysates, rotating at 4 °C, overnight. Samples were then washed for 10 min each with a series of 4 wash buffers that decreased in detergent concentration. Beads were then washed a final two times in × PBS, left in about 60 μl volume PBS, then frozen until ready for analysis by the mass spectrometry facility.

In preparation for mass spectrometry, proteins were reduced with 1.75 μl 15 mg/ml DTT in 10 mM TEAB, shaking at 56 °C for 50 min. Samples were then cooled to room temperature, the pH adjusted to 8 with 500 mM TEAB buffer, and alkylated with 1.8 μl 36 mg/ml iodoacetomide in 100 mM TEAB for 20 min at room temperature, in the dark. Next, 20 ng/μl trypsin (Promega) was added to proteolyze the samples at 37 °C, overnight. Supernatant was collected, and the beads were washed with 0.1× TFA 3 times, with washes added to supernatant. The pH was adjusted to acidic range, and peptides desalted on u-HLB Oasis plates, eluted with 60% acetonitrile/0.1% TFA, and dried. Ten per cent desalted peptides were analysed on a Nano LC-MS/MS instrument on Q Exactive Plus (Thermo) in FTFT mode. Tandem mass spectrometry data were searched with Mascot via PD2.2 against RefSeq2017_83 human species database and a small enzyme and standard (BSA)-containing database using the FilesRC option, with mass tolerance of 3 ppm on precursors and 0.01 Da on fragments, and annotating variable modifications such as oxidation on M, carbamidomethyl C, deamidation NQ, with and without biotin K. The Mascot .dat files were (1) compiled in Scaffold and (2) processed in PD2.2 to identify peptides and proteins using Percolator as a PSM validator.

Protein hits identified only in miniTurbo-TRIM37 BioID, and hits with spectral counts in miniTurbo-TRIM37 BioID that were twofold greater than those of mTurbo alone were considered as candidates for TRIM37 interaction. The filtered list of BioID hits was annotated with Gene Ontology (GO) terms via the Panther classification system[41] and analysed using the statistical overrepresentation test (binomial) to derive *P* values[42].

Antibody techniques

For immunoblot analyses, protein samples were separated by SDS–PAGE, transferred onto nitrocellulose membranes with a Trans-Blot Turbo Transfer System (BioRad) and then probed with the following primary antibodies: YL1/2 (rat anti-α-tubulin, ThermoFisher Scientific, MA1-80017, 1:3,000), TRIM37 (rabbit, Bethyl, A301-174A, 1:1,000), p53 (mouse, Dako, M7001, 1:1,000), β-actin (mouse, Sigma, A1978, 1:1,000), HA-11 (mouse, BioLegend, 901501, 1:1,000), GST (mouse, Sigma, G1160, 1:1,000), CEP192 (rabbit, home-made, 1:1,000), CDK5RAP2 (rabbit, Millipore, 06-1398, 1:2,500), pericentrin (rabbit, Abcam, ab4448, 1:2,500), cyclin A (mouse, SantaCruz Biotechnology, sc-53228, 1:1,000), phosphorylated histone H3 (rabbit, Millipore, 06-570, 1:2,000). Proteins were then detected using HRP-conjugated anti-mouse (goat, ThermoFisher Scientific, 31432, 1:1,000) or anti-rabbit (goat, ThermoFisher Scientific, 31462, 1:10000) secondary antibodies and enhanced chemiluminescence (Clarity, Bio-Rad). Signals were visualized and acquired using the Gel DocXRSystem (Bio-Rad).

For immunofluorescence, cells were grown on 18-mm glass coverslips and fixed for 10 min in either 4% formaldehyde at room temperature, or 100% ice-cold methanol at −20 °C for 10 min. Cells were blocked in 2.5% FBS, 200 mM glycine, and 0.1% Triton X-100 in PBS for 1 h. Antibody incubations were conducted in the blocking solution for 1 h. DNA was stained with DAPI and cells were mounted in ProLong Gold Antifade (Invitrogen). Staining was performed with the following primary antibodies: centrin (mouse, Millipore, 04-1624, 1:1,000), CDK5RAP2 (rabbit, Millipore, 06-1398, 1:2,500), γ-tubulin–Cy5 (directly labelled goat, raised against the following peptide: CDEYHAATRPDYISWGTQEQ, this study, 1:1,000), pericentrin (rabbit, Abcam, ab4448, 1:2,500), CEP192–Cy5 (directly labelled goat, raised against CEP192 amino acids 1–211, this study, 1:1,000), YL1/2 (rat anti-α-tubulin, ThermoFisher Scientific, MA1-80017, 1:3,000), EB1 (mouse, Santa Cruz, sc-47704, 1:200).

Immunofluorescence images were collected using a Deltavision Elite system (GE Healthcare) controlling a Scientific CMOS camera (pco.edge 5.5). Acquisition parameters were controlled by SoftWoRx suite (GE Healthcare). Images were collected at room temperature (25 °C) using an Olympus 40× 1.35 NA, 60× 1.42 NA or Olympus 100× 1.4 NA oil objective at 0.2-μm *z*-sections. Images were acquired using Applied Precision immersion oil (*N* = 1.516). For quantification of signal intensity at the centrosome, deconvolved 2D maximum intensity projections were saved as 16-bit TIFF images. Signal intensity was determined using ImageJ by drawing a circular region of interest (ROI) around the centriole (ROI S). A larger concentric circle (ROI L) was drawn around ROI S. ROI S (S) and ROI L (L) were transferred to the channel of interest and the signal in ROI S was calculated using the formula $I_S - [(I_L - I_S/A_L - A_S) \times A_S]$, in which A is area and I is integrated pixel intensity.

To measure the distance between two centrosomes in prophase, cells were fixed in 4% formaldehyde at room temperature for 10 min. Coverslips were blocked and stained as above with the following primary antibodies: CENP-F (sheep, a gift from S. Taylor, 1:2,000), phospho-histone H3 (mouse, Cell Signaling, 9701, 1:2,000), and CEP192 (rabbit, a gift from K. Oegema, this study, 1:2,000). Phospho-histone-H3-positive cells with a nuclear-envelope-localized CENP-F signal were selected for analysis. The distance between two centrosomes was measured from 3D-image stacks using Imaris (Bitplane) software.

Live cell microscopy

Fluorescent cell lines were seeded into either 4-chamber, 35-mm glass-bottom culture dishes (Greiner) or 4-well chamber slides (Ibidi) and maintained at 37 °C in an environmental control station. Long-term time-lapse imaging was performed using a Deltavision Elite system (GE Healthcare) controlling a Scientific CMOS camera (pco.edge 5.5.). Images were acquired with an Olympus 40× 1.4 NA oil objective. Every 5 min, 7 × 3-µm *z*-sections were acquired in respective fluorescent channels and by differential inference contrast. Time-lapse imaging of PCM foci and EB1 or EB3 comets was performed using a Lecia SP-8 confocal microscope, equipped with a resonance scanner, and 405-nm, 488-nm, 552-nm and 638-nm laser lines. Images were acquired with a Leica 40× 1.3 NA or 63× 1.4 NA oil objectives. For time-lapse imaging of PCM foci, images were captured every 5 min in 20 × 1-µm *z*-sections. For time-lapse imaging of EB1 or EB3 comets, images were collected every 2 s in a single *z*-plane. Movies were deconvolved using the LIGHTNING adaptive approach, and assembled and analysed in FIJI. Mitotic duration was calculated as the time taken from nuclear envelope breakdown to the onset of anaphase.

Microtubule regrowth assay

Cells were treated with 3.3 µM nocodazole for 1 h to disrupt the microtubule network, then quickly rinsed 3× with warmed medium (37 °C) to remove the drug. Cells were then incubated in warmed medium for 90 s to allow microtubule regrowth, fixed in 100% ice-cold methanol for 10 min, and processed as described in 'Antibody techniques' for immunofluorescence. For fluorescence intensity quantification, images were analysed in ImageJ using a circular area of 5 µm to quantify α-tubulin and EB1 signals around centrosomes. Background fluorescence using a circle of corresponding size was subtracted from each measurement.

PLK4i survival assays

For short-term survival assays, cells seeded in triplicate at 1.25×10^4 cells per well in 6-well plates were treated with either DMSO control or PLK4i (10 µM 3MB-PP1, or 125 nM centrinone) 16 h later. After the indicated number of days, cells were fixed and stained using 0.5% (w/v) crystal violet in 20% (v/v) methanol for 5 min. Excess of reagent was extensively washed with distilled water and plates dried overnight. For quantification, bound crystal violet was dissolved in 10% (v/v) acetic acid in dH_2O and absorbance of 1:50 dilutions were measured at 595 nm in a WPA S800 Spectrawave spectrophotometer (Biochrom). Optical density at 595 nm was measured as a quantitative metric of relative growth.

For long-term clonogenic survival assays, 500 cells were seeded in a 10-cm^2 culture dish in triplicate and left to adhere overnight. Cells were treated the next day and left to grow for about 14 d or until colonies were visible by eye. Plates were then stained with crystal violet dye (Sigma) and colonies counted. The relative colony growth was assessed relative to DMSO control plates.

RNA extraction and reverse-transcription qPCR

Total RNA was extracted using the RNeasy Plus Mini kit (Qiagen) and reverse transcription was performed using the iScript cDNA Synthesis Kit (Bio-Rad) following manufacturer's protocol. *TRIM37* transcripts were measured by qPCR in triplicate on a CFX96 Real-Time Analyzer (Bio-Rad) using Quantifast SYBR Green reagent (QIAGEN), normalized to reference gene *SMG9* and quantified using the $\Delta\Delta C_t$ method to obtain relative expression. Thermocycling conditions were set as follows: 1 cycle (95 °C for 5 mins), 40 cycles (95 °C for 15 s, 58 °C for 60 s). Primer sequences were *TRIM37* forward (5′-TCAGCTGTATTAGGCGCTGG-3′), *TRIM37* reverse (5′-ACTTCTTCTGCCCAACGACA-3′), *SMG9* forward (5′- GCCCTGGAGAAGAATGAA-3′) and *SMG9* reverse (5′-GGTGAAA GACAACAGCATC-3′).

Flow cytometry

G0/G1, S and G2/M cell cycle profiles were assessed using 5′-bromouridine (BrdU) incorporation and propidium iodide (PI) staining. Cells were pulsed with 10 µM BrdU (Sigma-Aldrich) for 30 min, trypsinized and washed with 1% BSA in PBS (1,500 rpm, 5 min) before being fixed in 70% ethanol. DNA denaturation was performed using a solution of 0.2 mg/ml of pepsin (Sigma-Aldrich) in 2 M HCl for 20 min at room temperature. Cells were washed twice with PBS and re-suspended in a solution containing anti-BrdU–FITC-conjugated antibody (rat, Bio-Rad, MCA2060FT, 1:100) in 0.5% (v/v) Tween-20, 0.5% (v/v) BSA in PBS and incubated for 1 h in the dark. For determination of the mitotic phase cell population (M), cells fixed in 70% ethanol were permeabilized with 0.2% Tween-20 in 2 M HCl for 10 min. Cells were then stained with anti-phosphorylated histone H3 (Ser10) antibody (mouse, Cell Signaling Technology, 9706, 1:50) in 1% (v/v) BSA in PBS for 3 h. Cells were then washed twice with PBS, re-suspended in a solution containing anti-mouse Alexa-Fluor-488-conjugated antibody (goat, ThermoFisher Scientific, A-11029, 1:250) in 1% (v/v) BSA in PBS and incubated for 1 h in the dark. For total DNA staining, including those used for determination of sub-G1 population and ploidy analyses, a 20-min incubation at 37 °C in a solution of PI/RNaseA (10 µg/ml and 0.1 mg/ml, respectively) in PBS was performed. Samples were analysed using an Attune NxT flow cytometer (Life Technologies) and data processing was done using FlowJo software.

SA-β-gal staining

The SA-β-gal activity of DMSO- or centrinone-treated MCF-7 cells was assessed using a staining kit (Cell Signaling, no. 9860), as per the manufacturer's protocol. Stained cells were imaged with a Nikon wide-field TE2000U Microscope at 200× magnification. For quantification, up to 200 cells per condition were counted across multiple fields to determine the percentage of SA-β-gal-positive cells.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Data that support the findings of this study are available from the corresponding authors upon reasonable request. Source data are provided with the paper.

34. Di, Z. et al. Ultra high content image analysis and phenotype profiling of 3D cultured micro-tissues. *PLoS ONE* **9**, e109688 (2014).
35. Booij, T. H. et al. Development of a 3D tissue culture-based high-content screening platform that uses phenotypic profiling to discriminate selective inhibitors of receptor tyrosine kinases. *J. Biomol. Screen*. **21**, 912–922 (2016).
36. Sandercock, A. M. et al. Identification of anti-tumour biologics using primary tumour models, 3-D phenotypic screening and image-based multi-parametric profiling. *Mol. Cancer* **14**, 147 (2015).
37. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
38. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
39. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
40. Firat-Karalar, E. N. & Stearns, T. in *Methods in Cell Biology* **vol. 129** (eds Basto, R. & Oegema, K.) 153–170 (Academic, 2015).

41. Mi, H. et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45**, D183–D189 (2017).
42. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protocols* **8**, 1551–1566 (2013).

Acknowledgements This work was supported by a Cancer Research UK Career Development Fellowship C52690/A19270 (to J.R.C.), and by National Institutes of Health grants R01GM114119 and R01GM133897, an American Cancer Society Scholar grant RSG-16-156-01-CCG, and an American Cancer Society Mission Boost Grant MBG-19-173-01-MBG (to A.J.H.). Z.Y.Y. is supported by the National Science Scholarship from A*STAR, Singapore. J.R.C. holds a Lister Institute Research Prize Fellowship. The Wellcome Centre for Human Genetics is supported by Wellcome grant 090532/Z/09/Z. The C.J.L. and A.N.J.T laboratories are funded by NC3Rs (NC/P001262/1), Breast Cancer Now funding to the Breast Cancer Now Toby Robins Research Centre (CTR-Q4-Y2), private donations to the ICR Development Office, and NHS funding to the NIHR Biomedical Research Centres at Guy's and St Thomas' NHS Foundation Trust and King's College London, and the Royal Marsden Hospital. We also thank R. Peat for breast cancer cell lines.

Author contributions Z.Y.Y. and B.G.L. designed, performed and analysed the majority of the experiments, and prepared the figures. M.-A.D., D.M. and C.M.G. assisted with associated cytogenetic experiments. K.H.Z. performed and analysed centrosomal intensity quantification experiments. L.T.E. generated the cell lines and analysed the movies for the data in Fig. 4b, c. P.M.S. created and performed the imaging of the DLD-1 cell line that expresses CEP192–mNeonGreen. T.P. analysed centrosome separation in 17q23-amplified and non-17q23-amplified cell lines. E.P. created the *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells that express EB3–mNeonGreen, TUBG1–TagRFP and H2B–iRFP. L.A.R. analysed PCM foci in mitotic MDA-MB-436 and DLD-1 cells. R.M., E.G.K., L.M.B., D.N. and S.H. conducted and analysed 3D tumour cell line and PDO experiments, under the supervision of A.N.J.T. and C.J.L. J.R.C. and A.J.H. conceived and co-supervised the study, designed experiments, and analysed the data. Z.Y.Y., B.G.L., J.R.C. and A.J.H. co-wrote the manuscript.

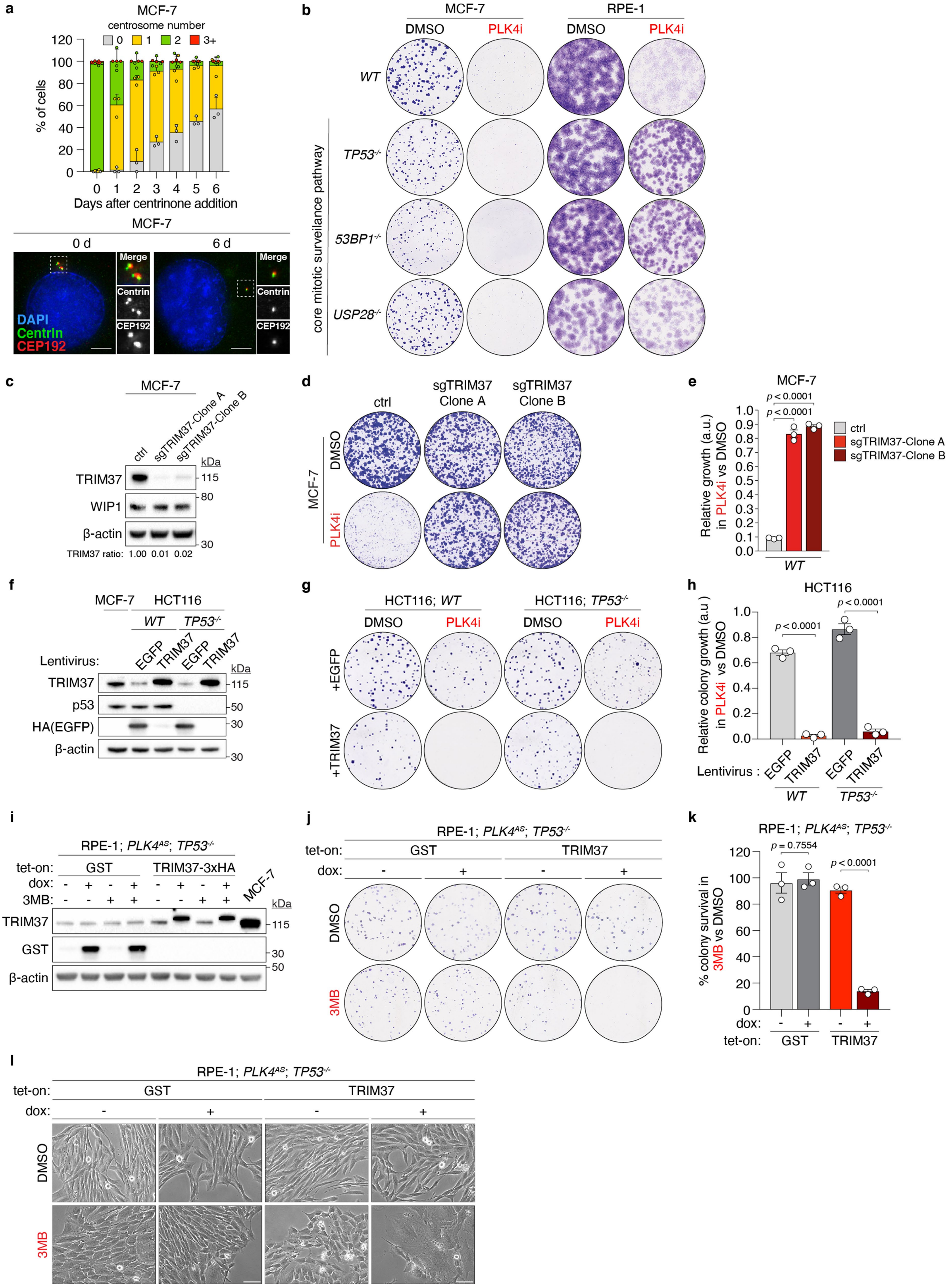
Competing interests C.J.L. received research funding from AstraZeneca, Merck KGaA and Artios; received consultancy, SAB membership or honoraria payments from Syncona, Sun Pharma, Gerson Lehrman Group, Merck KGaA, Vertex, AstraZeneca, Tango, 3rd Rock, Ono Pharma and Artios; and has stock in Tango and Ovibio. None of these is directly relevant to the work published and/or discussed in this Article. C.J.L. is also a named inventor on patents describing the use of DNA repair inhibitors and stands to gain from the development as part of the ICR Rewards to Inventors scheme. A.N.J.T received research funding from AstraZeneca and Merck KGaA that is not directly relevant to work published or discussed in this Article. A.N.J.T. also receives payments from the ICR Rewards to Inventors scheme associated with patents describing the use of DNA repair inhibitors and stands to gain from their development.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2690-1.

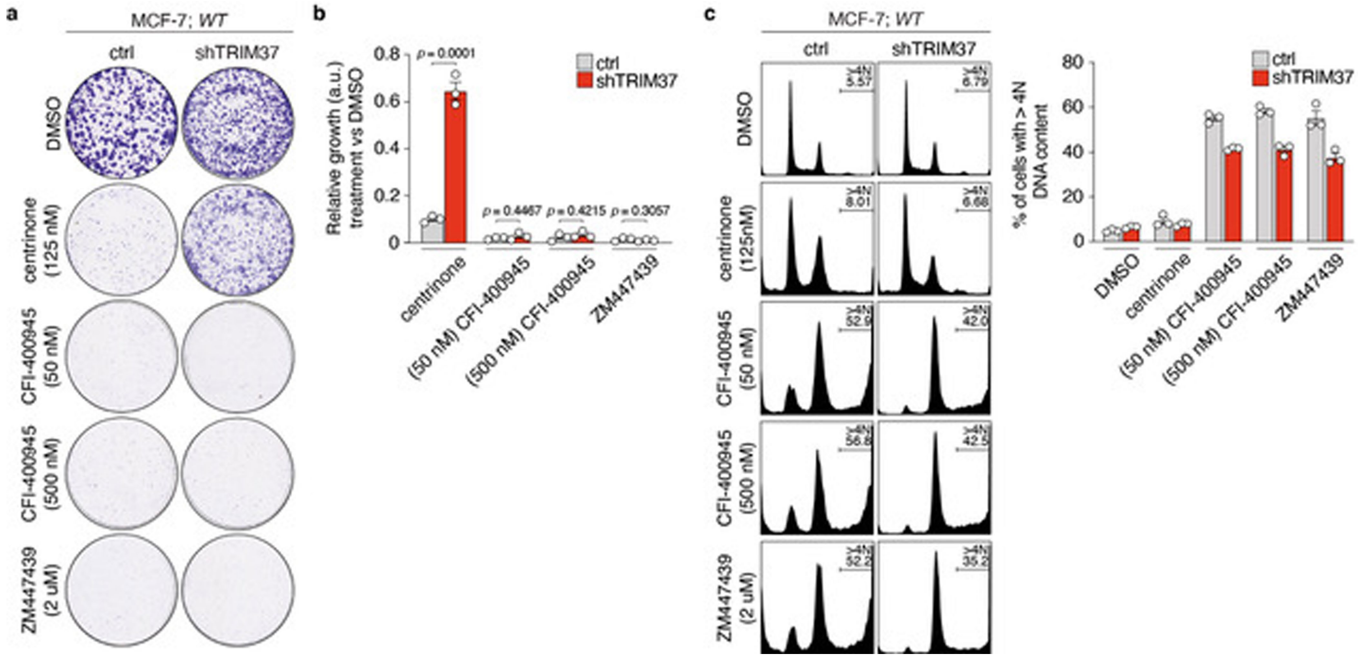
Correspondence and requests for materials should be addressed to J.R.C. or A.J.H.

Peer review information *Nature* thanks Renata Basto, Sarah McClelland and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

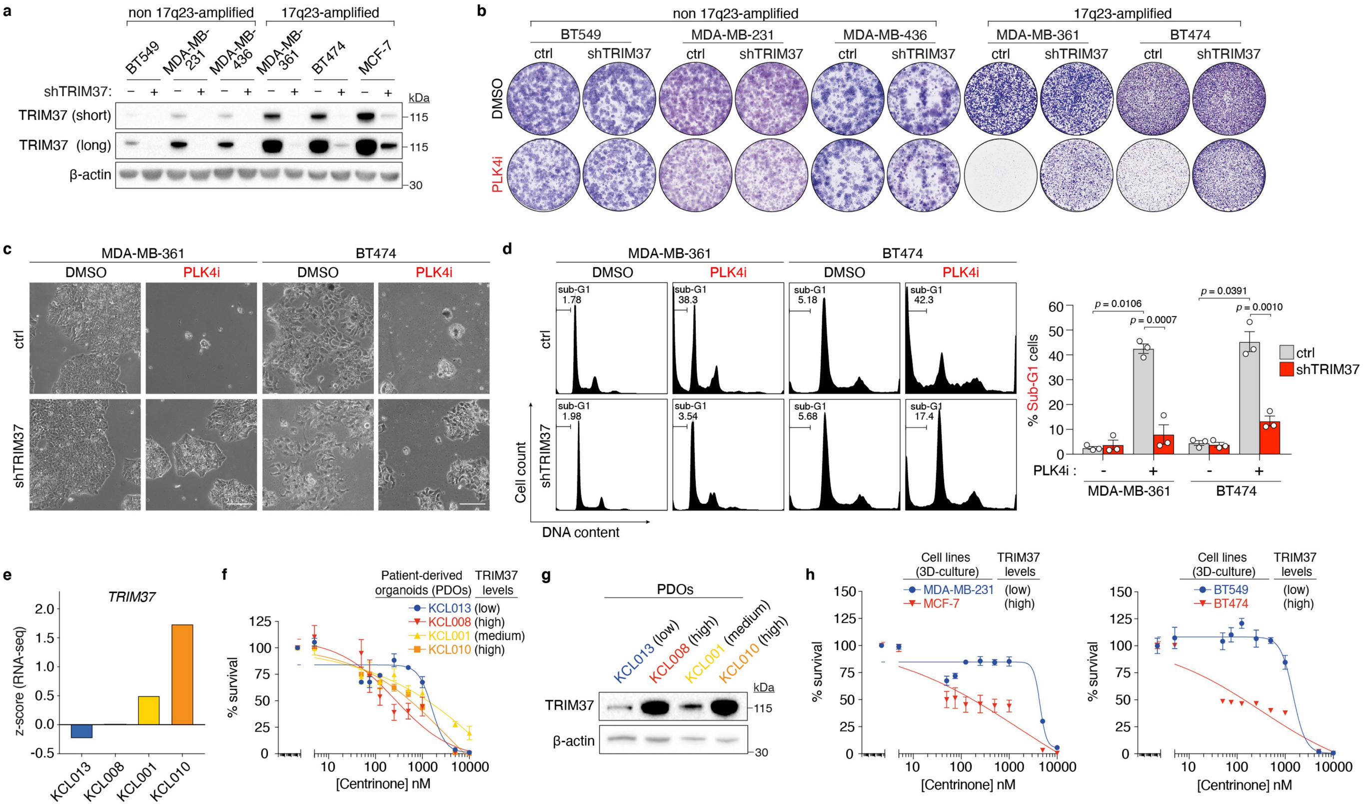


Extended Data Fig. 1 | See next page for caption.

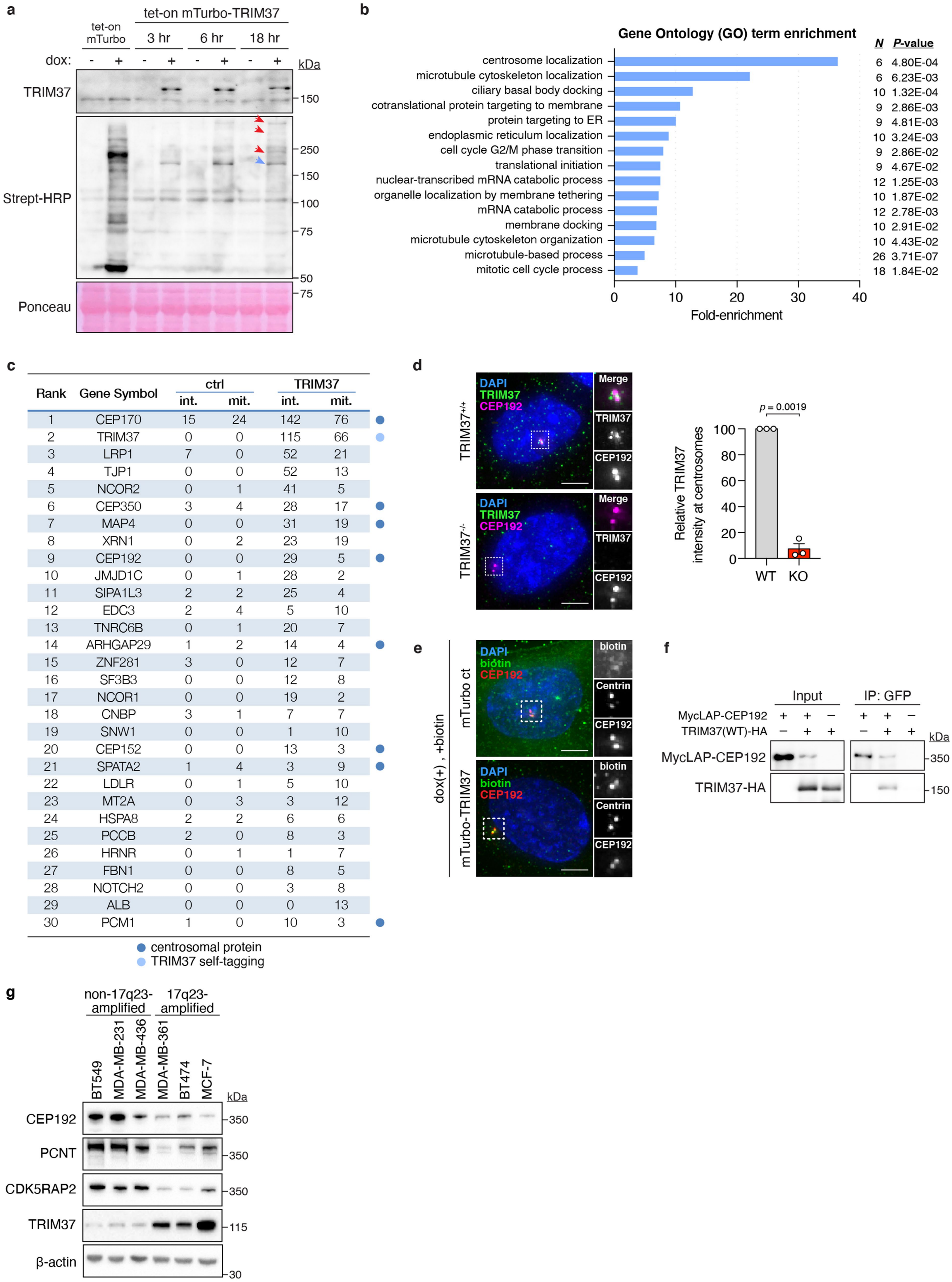
Extended Data Fig. 1 | TRIM37 overexpression in HCT116 and RPE-1 cells recapitulates synthetic lethality with centrosome loss. Related to Fig. 1. **a**, Top, centrosome number distribution in interphase MCF-7 cells at various times after addition of centrinone (PLK4i) (125 nM). Mean ± s.e.m. Bottom, representative images of centrosome staining (centrioles labelled by centrin, and PCM labelled by CEP192). $n = 3$ biological replicates, each comprising >100 cells. **b**, Representative data of a 14-day clonogenic survival assay of MCF-7 and RPE-1 cells with the indicated genotypes treated with DMSO (control) or centrinone (PLK4i) (125 nM). $n = 3$ biological replicates. **c**, Immunoblot showing TRIM37 protein levels in two WT MCF-7 clones stably expressing control vector or a *TRIM37*-targeting sgRNA. β-Actin, loading control. Representative data; $n = 3$ biological replicates. For gel source data, see Supplementary Fig. 1. **d**, Representative data of a 10-day clonogenic survival of indicated MCF-7 cell lines treated with DMSO (control) or centrinone (PLK4i) (125 nM). **e**, Quantification of $n = 3$ biological replicates in **d**. P values, unpaired two-tailed t-test. Mean ± s.e.m. **f**, Immunoblot of lysates prepared from WT and $TP53^{-/-}$ HCT116 cells expressing a control (*eGFP*) or *TRIM37* transgene. MCF-7 cells were used as a reference for TRIM37 protein overexpression in a 17q23-amplified cell line. β-Actin, loading control. Representative data; $n = 3$ biological replicates. For gel source data, see Supplementary Fig. 1. **g**, Representative data of a 14-day clonogenic survival assay of HCT116 cells treated with DMSO (control) or centrinone (PLK4i) (125 nM). **h**, Quantification of $n = 3$ biological replicates in **g**. P values, unpaired two-tailed t-test. Mean ± s.e.m. **i**, Immunoblot showing doxycycline-induced *GST* or *TRIM37* expression in $PLK4^{AS}$ $TP53^{-/-}$ RPE-1 cells. β-Actin, loading control. Representative data; $n = 3$ biological replicates. For gel source data, see Supplementary Fig. 1. **j**, Representative data of a 14-day colony survival assay of $PLK4^{AS}$ $TP53^{-/-}$ RPE-1 cells expressing doxycycline-inducible *GST* (control) or *TRIM37* transgenes, treated with DMSO (control) or 3MB-PP1 (3MB). AS, analogue sensitive. **k**, Quantification of $n = 3$ biological replicates in **j**. P values, unpaired two-tailed t-test. Mean ± s.e.m. **l**, Representative images of $PLK4^{AS}$ $TP53^{-/-}$ RPE-1 cells in **j**. Scale bars, 100 µm.



Extended Data Fig. 2 | Inhibitor selectivity for PLK4—and not other kinases—is required for the synthetic lethal killing of cells overexpressing TRIM37. a, Representative data of a 10-day clonogenic survival of indicated MCF-7 cell lines treated with DMSO (control), centrinone, CFI-400945 or ZM447439. Data acquired in parallel to experiment in Fig. 1c, d. **b**, Quantification of **a**, *n* = 3 biological replicates. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **c**, Left, representative flow cytometric analysis of DNA content in MCF-7 cells treated with DMSO (control), centrinone, CFI-400945 or ZM447439 for 3 d. Right, quantification of the percentage of cells with >4N DNA content (polyploidy). *n* = 3 biological replicates. Mean ± s.e.m.



Extended Data Fig. 3 | Additional characterization of TRIM37 expression and synthetic lethality in breast cancer cell lines and patient-derived organoids (PDOs). Related to Fig. 1. **a**, Immunoblot showing TRIM37 protein levels in the indicated 17q23-amplified cell lines (MDA-MB-361, BT474 and MCF-7) and non-17q23-amplified cell lines (BT549, MDA-MB-231 and MDA-MB-436) expressing control or *TRIM37*-targeting shRNA. β-Actin, loading control. Representative data; *n* = 3 biological replicates. For gel source data, see Supplementary Fig. 1. **b**, Clonogenic survival of 17q23-amplified and non-17q23-amplified cell lines treated with DMSO (control) or centrinone (PLK4i) (125 nM). Representative data; *n* = 3 biological replicates. **c**, Images of DMSO- or PLK4i-treated MDA-MB-361 and BT474 cells expressing control vector or *TRIM37*-targeting shRNA. Scale bars, 200 μm. Representative data; *n* = 3 biological replicates. **d**, Left, Representative flow cytometric DNA content analysis in DMSO- or PLK4i-treated MDA-MB-361 and BT474 cells. Percentages of sub-G1 events are indicated. Right, percentage of sub-G1 cells across *n* = 3 biological replicates. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **e**, *TRIM37* gene expression in PDOs. Gene expression is reported as a *z*-score derived from RNA-seq data sets across *n* = 22 independent biological samples. **f**, Viability of patient-derived breast tumour organoids following a 14-d exposure to the indicated concentrations of centrinone. Data from *n* = 2 biological replicates are shown. Mean ± s.e.m. **g**, Immunoblot showing TRIM37 protein levels in PDOs. β-Actin, loading control. Data from *n* = 1 biological replicate. For gel source data, see Supplementary Fig. 1. **h**, Viability of 3D cultures of the indicated cell lines following a 14-d exposure to the indicated concentrations of centrinone. Left, *n* = 2 biological replicates, Mean ± s.e.m. Right, *n* = 4 technical replicates, Mean ± s.e.m.



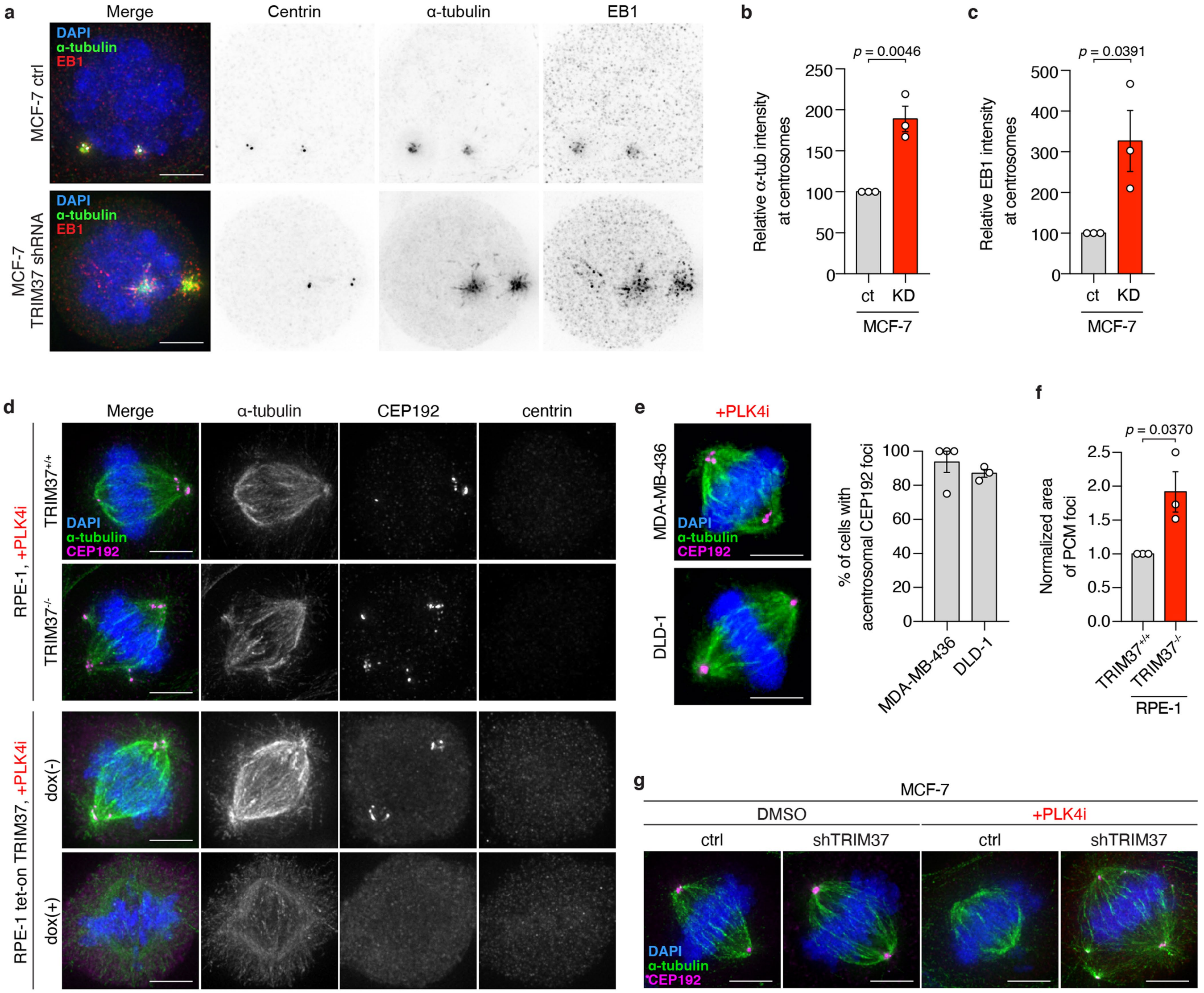
c

Rank	Gene Symbol	ctrl		TRIM37		
		int.	mit.	int.	mit.	
1	CEP170	15	24	142	76	●
2	TRIM37	0	0	115	66	○
3	LRP1	7	0	52	21	
4	TJP1	0	0	52	13	
5	NCOR2	0	1	41	5	
6	CEP350	3	4	28	17	●
7	MAP4	0	0	31	19	●
8	XRN1	0	2	23	19	
9	CEP192	0	0	29	5	●
10	JMJD1C	0	1	28	2	
11	SIPA1L3	2	2	25	4	
12	EDC3	2	4	5	10	
13	TNRC6B	0	1	20	7	
14	ARHGAP29	1	2	14	4	●
15	ZNF281	3	0	12	7	
16	SF3B3	0	0	12	8	
17	NCOR1	0	0	19	2	
18	CNBP	3	1	7	7	
19	SNW1	0	0	1	10	
20	CEP152	0	0	13	3	●
21	SPATA2	1	4	3	9	●
22	LDLR	0	1	5	10	
23	MT2A	0	3	3	12	
24	HSPA8	2	2	6	6	
25	PCCB	2	0	8	3	
26	HRNR	0	1	1	7	
27	FBN1	0	0	8	5	
28	NOTCH2	0	0	3	8	
29	ALB	0	0	0	13	
30	PCM1	1	0	10	3	●

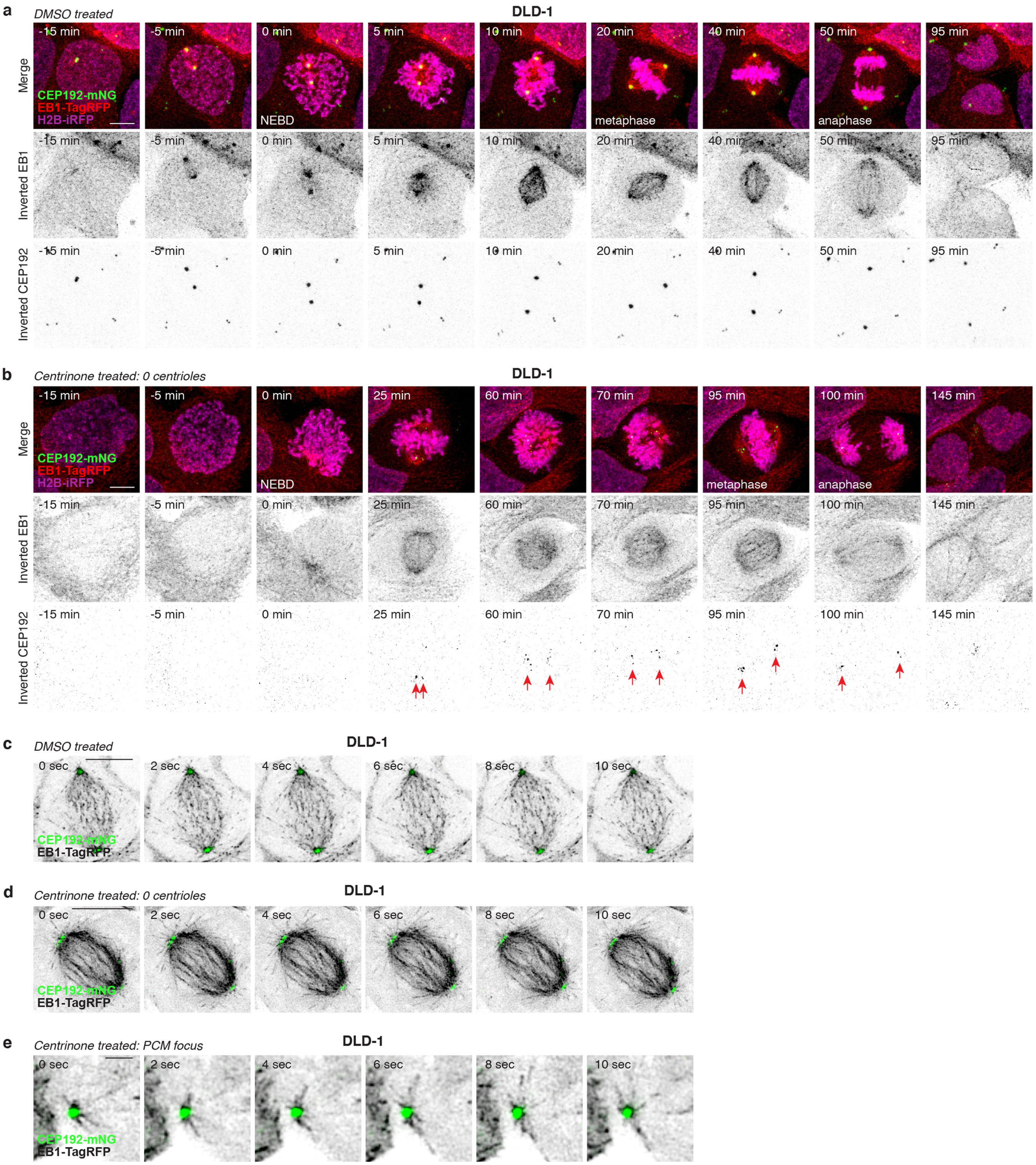
● centrosomal protein
○ TRIM37 self-tagging

Extended Data Fig. 4 | See next page for caption.

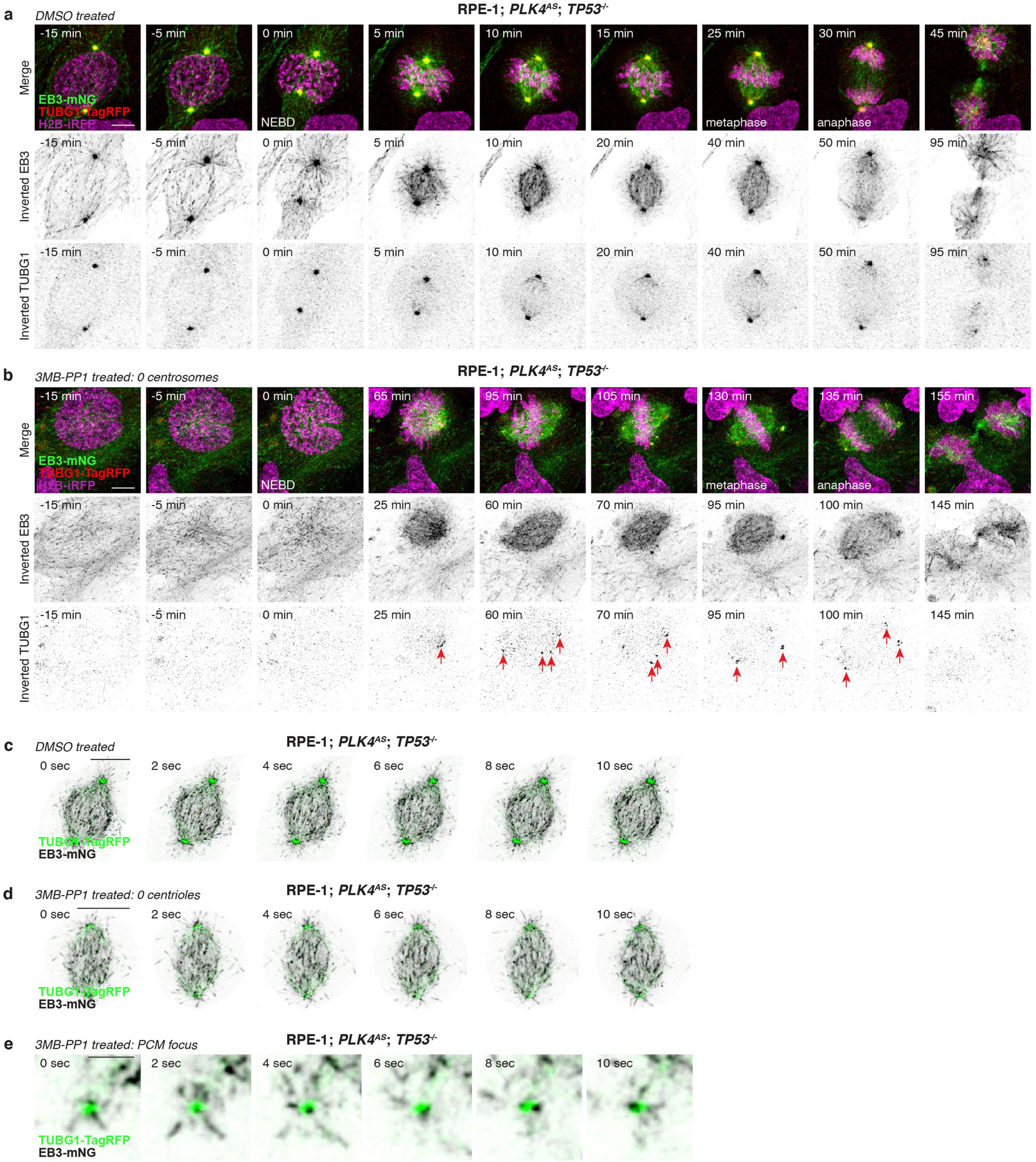
Extended Data Fig. 4 | TRIM37 localizes to centrosomes, where it interacts with, and regulates the abundance of PCM proteins. Related to Fig. 3. **a**, Immunoblot showing TRIM37 and biotinylated proximity interactors. Ponceau-stained blot indicates loading. Data are from a single experiment performed in duplicate. For gel source data, see Supplementary Fig. 1. **b**, Gene ontology analysis of mass spectrometry data. **c**, Thresholded mass spectrometry results displaying the top 30 proximity interactors by spectral count. Interactors were filtered to isolate those with >2 × more peptides in the mTurbo-TRIM37 sample compared to control. **d**, Left, immunofluorescence of TRIM37 in *TRIM37*$^{+/+}$ and *TRIM37*$^{-/-}$ RPE-1 cells. Scale bars, 5 µm. Right, quantification of TRIM37 intensity at the centrosome in RPE-1 cells. $n = 3$ biological replicates, each comprising >40 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **e**, Immunofluorescence of biotin-labelled proteins in mTurbo cell lines. Representative data; $n = 3$ biological replicates. Scale bars, 5 µm. **f**, Co-immunoprecipitation showing the interaction of TRIM37 with CEP192. Representative data; $n = 3$ biological replicates. For gel source data, see Supplementary Fig. 1. **g**, Immunoblot showing the levels of TRIM37 and PCM components in non-17q23-amplified versus 17q23-amplified cell lines. β-Actin, loading control. Representative data; $n = 3$ biological replicates. For gel source data, see Supplementary Fig. 1.



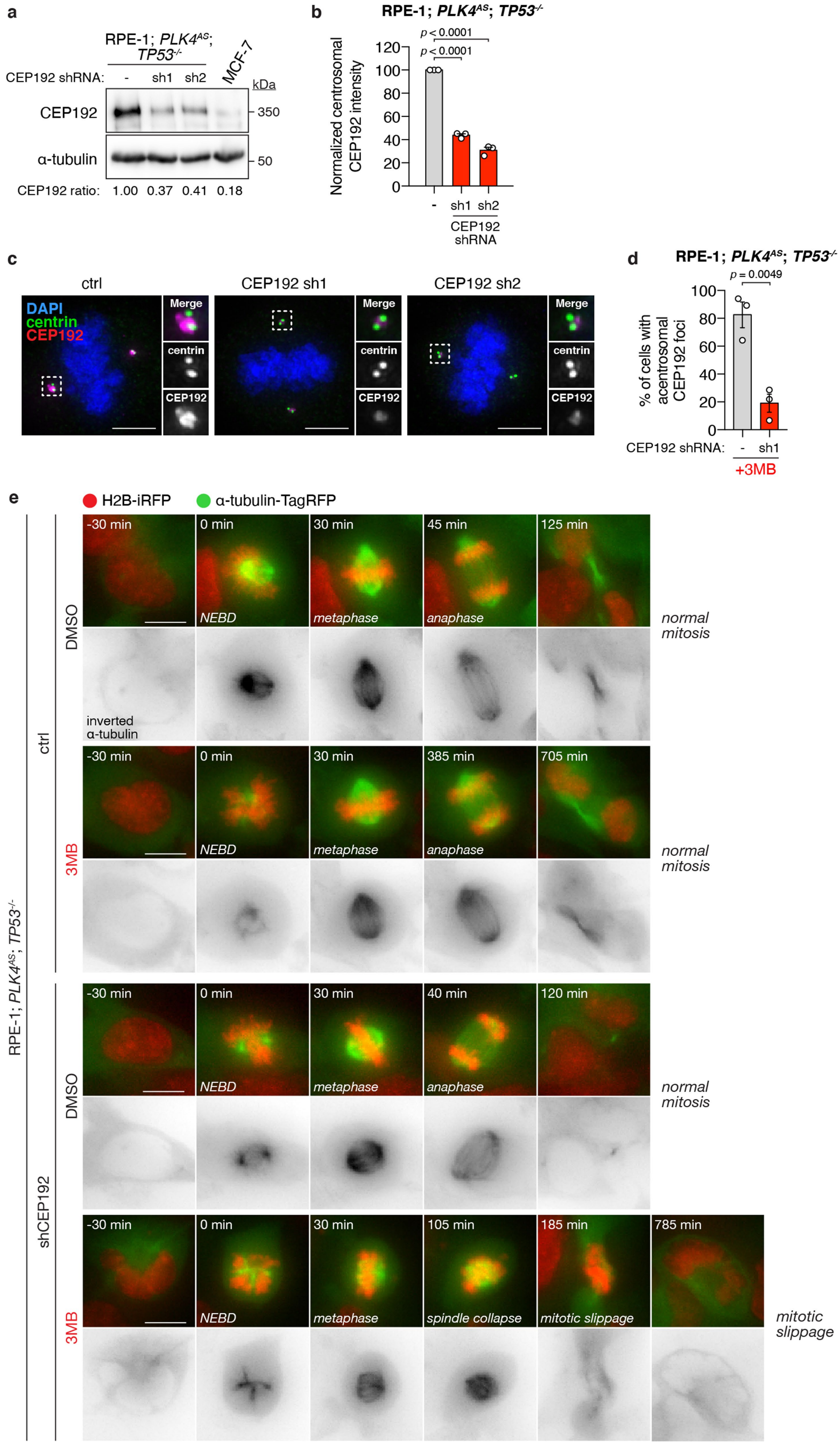
Extended Data Fig. 5 | TRIM37 suppresses microtubule nucleation by the centrosome and supresses the formation of non-centrosomal PCM foci. Related to Fig. 3. **a**, Microtubule regrowth following nocodazole washout in control vector- or *TRIM37*-shRNA-expressing MCF-7 mitotic cells. Representative images from **b**. $n = 3$ biological replicates. Scale bars, 5 µm. **b**, Quantification of microtubule regrowth following nocodazole washout in control vector- or *TRIM37*-shRNA-expressing MCF-7 mitotic cells. $n = 3$ biological replicates, each with >25 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **c**, Quantification of centrosomal EB1 intensity following nocodazole washout in control vector- or *TRIM37*-shRNA-expressing MCF-7 mitotic cells. $n = 3$ biological replicates, each with >25 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **d**, Representative images of mitotic PCM foci in acentrosomal RPE-1 cells described in Fig. 3d. $n = 3$ biological replicates. Scale bars, 5 µm. **e**, Left, representative images of mitotic PCM foci in acentrosomal MDA-MB-436 and DLD-1 cells. Scale bars, 5 µm. Right, quantification of mitotic PCM foci in centrinone-treated MDA-MB-436 and DLD-1 cells that lacked centrosomes. $n \geq 3$ biological replicates, each comprising ≥ 84 cells for DLD-1 cells and ≥ 6 cells for MDA-MB-436 cells. Mean ± s.e.m. **f**, Quantification of CEP192 foci area in *TRIM37*$^{+/+}$ versus *TRIM37*$^{-/-}$ RPE-1 cells in **d**. $n = 3$ biological replicates, each comprising >20 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **g**, Representative images for spindle length analysis in indicated MCF-7 cells described in Fig. 3g. $n = 3$ biological replicates. Scale bars, 5 µm.



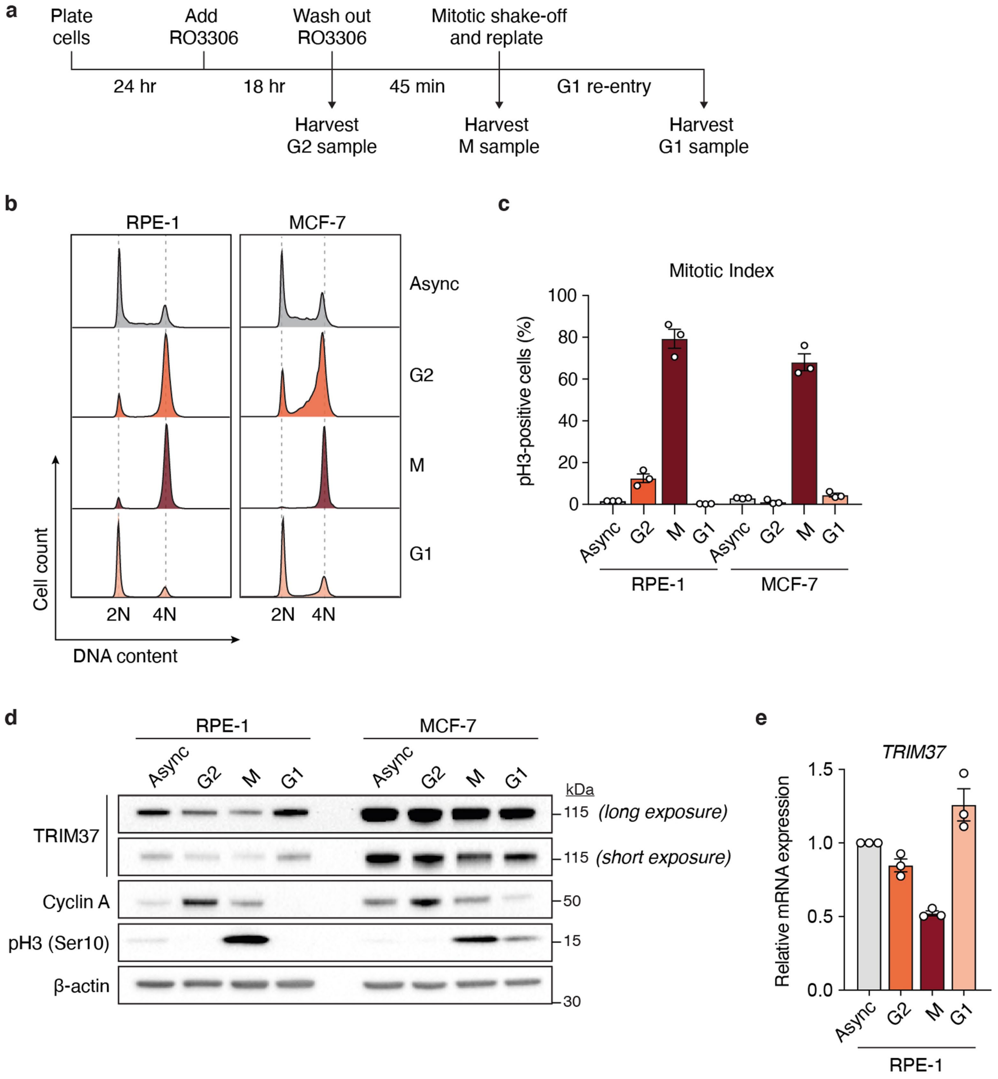
Extended Data Fig. 6 | Non-centrosomal PCM foci nucleate microtubules and contribute to spindle assembly in DLD-1 cells. Related to Fig. 3. **a**, Representative time-lapse images of mitosis in DMSO-treated control DLD-1 cells. *n* = 3 biological replicates. Scale bar, 5 µm. **b**, Representative time-lapse images of PCM foci formation during mitosis in acentrosomal DLD-1 cells. Scale bar, 5 µm. *n* = 3 biological replicates. Arrows indicate PCM foci. **c**, Representative time-lapse images of microtubule nucleation from centrosomes in the mitotic spindle of DMSO-treated control DLD-1 cells. *n* = 3 biological replicates. Scale bar, 5 µm. **d**, Representative time-lapse images of microtubule nucleation from PCM foci incorporated into the mitotic spindle of acentrosomal DLD-1 cells. *n* = 3 biological replicates. Scale bar, 5 µm. **e**, Representative time-lapse images of microtubule nucleation from a PCM focus before its incorporation into the mitotic spindle of the acentrosomal cell shown in **d**. *n* = 3 biological replicates. Scale bar, 1 µm.



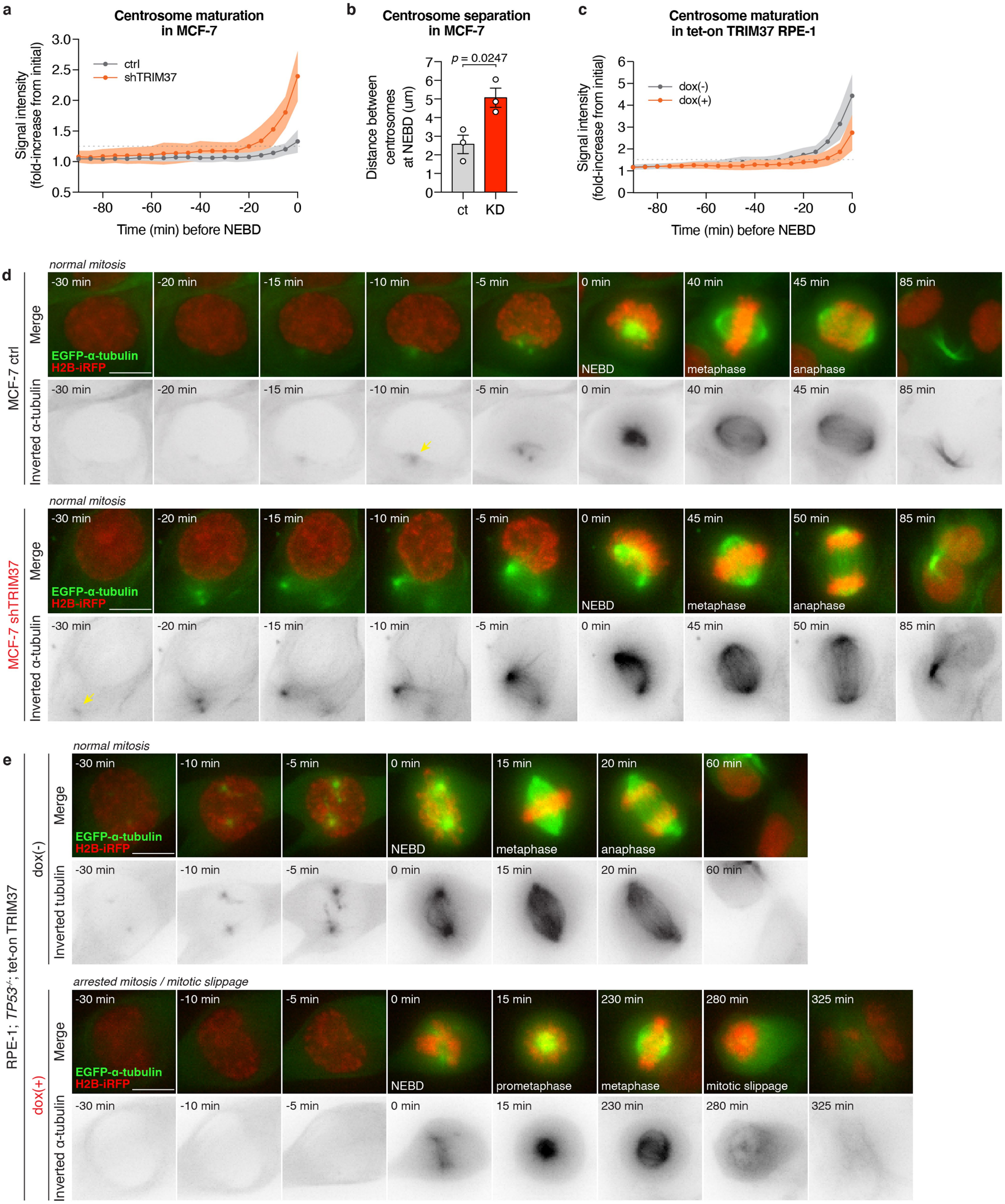
Extended Data Fig. 7 | Non-centrosomal PCM foci nucleate microtubules and contribute to spindle assembly in RPE-1 cells. Related to Fig. 3. **a**, Representative time-lapse images of mitosis in DMSO-treated control *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells. *n* = 3 biological replicates. Scale bar, 5 µm. **b**, Representative time-lapse images of PCM foci formation during mitosis in acentrosomal *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells. *n* = 3 biological replicates. Scale bar, 5 µm. Arrows indicate PCM foci. **c**, Representative time-lapse images of microtubule nucleation from centrosomes in the mitotic spindle of DMSO-treated control *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells. *n* = 3 biological replicates. Scale bars, 5 µm. **d**, Representative time-lapse images of microtubule nucleation from PCM foci incorporated into the mitotic spindle of acentrosomal *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells. *n* = 3 biological replicates. Scale bar, 5 µm. **e**, Representative time-lapse images of microtubule nucleation from a PCM focus before its incorporation into the mitotic spindle in acentrosomal *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1. *n* = 3 biological replicates. Scale bar, 1 µm.



Extended Data Fig. 8 | Depletion of CEP192 in RPE-1 cells recapitulates the synthetic lethal mitotic phenotypes observed in high-TRIM37 expressing cells. Related to Fig. 3. **a**, Immunoblot showing the CEP192 levels in indicated control and CEP192-depleted *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells. α-Tubulin, loading control. For gel source data, see Supplementary Fig. 1. **b**, Quantification of mitotic centrosomal CEP192 signal in the same cells as described in **a**. *n* = 3 biological replicates, each comprising >30 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **c**, Representative images of centrosomal CEP192 in the same cells as described in **a**. Scale bars, 5 µm. **d**, Quantification of the percentage of 3MB-PP1 (3MB)-treated *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells with acentrosomal mitotic CEP192 PCM foci. *n* = 3 biological replicates, each comprising >30 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **e**, Representative time-lapse images of mitotic progression in DMSO- or 3MB-PP1 (3MB)-treated control and CEP192-depleted *PLK4*^{AS} *TP53*$^{-/-}$ RPE-1 cells. Cells are labelled with H2B-iRFP and tagRFP-tubulin. *n* = 3 biological replicates.



Extended Data Fig. 9 | Cell cycle regulation of *TRIM37* expression. Related to Fig. 4. **a**, Schematic of the experimental protocol used for cell cycle synchronization. Samples were subjected to dual flow cytometry staining of phospho-histone Ser10 (pH3) to mark mitotic cells and propidium iodide (PI) to determine synchronization efficiency. M, mitotic phase. **b**, Flow cytometric DNA content analysis of samples collected according to **a**. Left, RPE-1. Right, MCF-7. Async, asynchronous. **c**, Mitotic index of cell cycle samples as determined by the percentage of pH3-positive cells with 4N DNA content. *n* = 3 biological replicates. Mean ± s.e.m. **d**, Immunoblot showing endogenous TRIM37, cyclin A and pH3 in samples analysed in **b**. β-Actin, loading control. For gel source data, see Supplementary Fig. 1. **e**, RT-qPCR analysis indicating relative *TRIM37* mRNA expression in RPE-1 cells analysed in **b**. Data were normalized to *TRIM37* mRNA expression in asynchronous cells. *n* = 3 biological replicates. Mean ± s.e.m.



Extended Data Fig. 10 | TRIM37 overexpression delays centrosome maturation in G2/M phase. Related to Fig. 4. **a**, Quantification of centrosomal α-tubulin intensity from time-lapse movies of dividing MCF-7 cells expressing either control vector or *TRIM37*-targeting shRNA. Quantification of >20 cells. Mean ± s.e.m. **b**, Quantification of the distance between the two centrosomes at NEBD in MCF-7 cells. $n = 3$ biological replicates, each comprising >12 cells. *P* values, unpaired two-tailed *t*-test. Mean ± s.e.m. **c**, Quantification of centrosomal α-tubulin intensity from time-lapse movies of dividing RPE-1 tet-on TRIM37 cells. Quantification of >20 cells. Mean ± s.e.m. **d**, Representative time-lapse images of centrosome maturation in MCF-7 cells. $n = 3$ biological replicates. Scale bars, 5 μm. **e**, Representative time-lapse images of centrosome maturation in RPE-1 tet-on TRIM37 cells. $n = 3$ biological replicates. Scale bars, 5 μm.

nature research

Corresponding author(s): J. Ross Chapman, Ph.D
Andrew J. Holland, Ph.D

Last updated by author(s): May 6, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Gel/membrane Imaging: Image Lab v5.2.1; Flow cytometry: Attune NxT Software V2.5; RT-PCR: Bio-Rad CFX Manager™ Software Immunofluorescence: GE Healthcare Deltavision Elite system and SoftWoRx suite or Leica Microsystems and LAS X elements.
Data analysis	FlowJo v10; GraphPad Prism v7 was typically used for all presented statistical analyses; Image analysis was preformed using FIJI or Imaris v9.2.1 (Bitplane).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All source data for graphs and gels in Fig. 1–4 and Extended Data Fig. 1–10 are available as .xslx tables and Supplementary Information within the manuscript. Other data that support the findings of this study are available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine the experimental sample size. All experiments conducted with cell lines were performed with multiple biological replicates based on previous experience with the sample size required to identify statistically significant effect sizes.
Data exclusions	No data were excluded from the analyses.
Replication	Following extensive optimization, biological experiments were typically performed in 3 biological replicates (each performed identically on different days) with consistent results. In the some cases, each biological replicate (e.g. clonogenic assay, flow cytometric sample) involved 2-3 technical replicates. Attempts to replicate findings were successful for all of the experiments presented in the manuscript.
Randomization	Experiments were performed using populations of manipulated cell lines and therefore randomization was not appropriate.
Blinding	Investigators were not blinded to the experimental conditions used during most experiments. The data reported are not subjective but rather based on quantitative analysis of phenotypes such as cell survival, distance, mitotic time and error frequency.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☐	☒	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☐	☒	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used

Antibodies used in western blot studies:

Primary
Rb anti TRIM37: Bethyl, A301-174A, 1:1000 (Holland, Chapman)
M anti p53: Dako, M7001, 1:1000 (Chapman)
M anti b-actin :Sigma, A1978, 1:1000 (Chapman)
Rb anti 53BP1: Novus Biologicals, NB100-304, 1:2000 (Chapman)
M anti HA-11: BioLegend, 901501, 1:1000 (Chapman)
M anti GST: Sigma, G1160, 1:1000 (Chapman)
Rb anti CEP192: home-made, 1:1000
Rb anti CDK5RAP2: Millipore, 06-1398, 1:2500 (Holland, Chapman)
Rb anti Pericentrin: Abcam, ab4448, 1:2500 (Holland, Chapman)
M anti cyclin A: SantaCruz Biotechnology, sc-53228, 1:1000 (Chapman)
Rb anti phosphorylated Histone H3: Millipore, 06-570, 1:2000 (Chapman)

Secondary
G anti-mouse HRP-conjugated: ThermoFisher Scientific, 31432, 1:1000 (Chapman)
G anti-rabbit HRP-conjugated: ThermoFisher Scientific, 31462, 1:10000 (Chapman)

Antibodies used in Immunofluorescence studies:

M anti Centrin: Millipore, 04-1624, 1:1000 (Holland)

Rb anti CDK5RAP2: Millipore, 06-1398, 1:2500 (Holland)
G anti g-tubulin-Cy5: raised against the following peptide: CDEYHAATRPDYISWGTQEQ, home made, 1:1000
Rb anti Pericentrin: Abcam, ab4448, 1:2500 (Holland)
G anti CEP192-Cy5: raised against CEP192 a.a. 1-211, home made, 1:1000
R anti a-tubulin: ThermoFisher Scientific, MA1-80017, 1:3000 (Holland)
M anti EB1: Santa Cruz, sc-47704, 1:200 (Holland)
S anti CENP-F: a gift from Stephen Taylor at the University of Manchester, 1:2000
M anti phosphorylated Histone H3: Cell Signaling, 9701, 1:2000 (Holland)
Rb anti CEP192: a gift from Karen Oegema at the University of California at San Diego, this study, 1:2000

Antibodies used in flow cytometry studies:

Primary
R anti BrdU-FITC: BioRad, MCA2060FT, 1:100
M anti phosphorylated Histone H3 (Ser10): Cell Signaling, 9706, 1:50

Secondary
G anti-mouse Alexa Fluor 488-conjugated: ThermoFisher Scientific, A-11029, 1:250)

Validation

All of our homemade antibodies were validated by immunoblotting and immunofluorescence to ensure the loss of signal after RNAi depletion or CRISPR/Cas9 knockout of the target protein. When available, we purchased commercial antibodies that have been previously validated in multiple independent studies. Validation procedures used for commercial antibodies are described by the respective manufacturers. In cases where this was not possible, commercial antibodies were validated in house in the same way we validate our homemade antibodies.

All antibodies used in flow cytometry studies were validated by the manufacturers as suitable for use in flow cytometry assays against specific antigens/markers.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

MCF-7 cell lines (WT, TP53BP1-/-, USP28-/-, and TP53-/-) were available in our lab and have been previously described (Cuella-Martin et al., 2016).
HEK293FT, MDA-MB-231, MDA-MB-436, MDA-MB-361 and BT474 cell lines were obtained from the Francis Crick Institute Cell Services.
HCT116 cell line was a gift from Ian Tomlinson.
RPE-1 cell lines, specifically the PLK4AS; TP53-/-; RPE-1 cell line were available in our lab and have been previously described (Lambrus et al., 2016).
DLD-1 and BT549 cell lines were obtained from Stephen Taylor, University of Manchester and Saraswati Sukumar, Johns Hopkins School of Medicine, respectively.

Authentication

MCF-7 cell lines (WT, TP53BP1-/-, USP28-/-, and TP53-/-) have been previously validated (Cuella-Martin et al., 2016), and shRNA expressing lines have been additionally validated by STR profiling.
MDA-MB-231, MDA-MB-436, MDA-MB-361 and BT474 cell lines were validated by STR profiling by the Francis Crick Institute Cell Services.
RPE-1 (WT, TP53BP1-/-, USP28-/-, and TP53-/-) cell lines were validated by western blotting (Supplementary Figure 2) and have been additionally validated by STR profiling.
PLK4AS; TP53-/-; RPE-1 cell line has been previously validated (Lambrus et al., 2016) and have been additionally validated by STR profiling.
HCT116 cell line TP53+/+ and TP53-/- statuses were further validated by western blotting (Extended Data Fig. 1d).
DLD-1 and BT549 cell lines were validated by STR profiling.
HEK293FT cells were used as a packaging cell line for lentiviral production, respectively, and were not further authenticated.

Mycoplasma contamination

Yes - we maintain a very strict regime of mycoplasma testing, and no cell-line tested positive.

Commonly misidentified lines
(See ICLAC register)

We have checked the ICLAC register and the cell lines used in our studies are not on the list of misidentified cell lines.

Human research participants

Policy information about studies involving human research participants

Population characteristics

Adult female patients with breast cancer as part of the BTBC study (see below). This study was limited to working with human tissue samples and data.

Recruitment

Human breast tumour samples were obtained from adult female patients after informed consent as part of a non-interventional clinical trial, the BTBC study, described below.

Ethics oversight

BTBC study: UK National Research Ethics Service - Research Ethics Committee London - (BTBC study REC no.: 13/LO/1248, IRAS

Ethics oversight	ID 131133; Principal Investigator: Prof. Andrew Tutt; Study Title: “Analysis of functional immune cell stroma and malignant cell interactions in breast cancer in order to discover and develop diagnostics and therapies in breast cancer subtypes”).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cell lines were used for all flow cytometric studies and sourced as indicated above. For sample preparation, cells were trypsinised and washed with 1% BSA in PBS before being fixed in 70% ethanol. Cells were then subjected to staining as described in the methods section.
Instrument	Samples were acquired on an Attune NxT (Life Technologies).
Software	Samples were analysed using FlowJo v10 (Tree Star).
Cell population abundance	Cell sorting was not necessary to evaluate sub-G1 events, cell-cycle phases and ploidy. Thus, it was not performed for these assays.
Gating strategy	Sub-G1/ ploidy analyses: Fixed cells were first gated to exclude debris (FSC-A vs SSC-A), then gated to select for singlets (FSC-H vs FSC-A), and finally assessed for DNA content by Propidium Iodide staining. Cell cycle/mitotic index analyses: Fixed cells were first gated to exclude debris (FSC-A vs SSC-A), then gated to select for singlets (FSC-H vs FSC-A), and finally assessed for DNA content by Propidium Iodide with BrdU or pH3 (negative or negative) stained fractions , when required. Please refer to (Supplementary Figure 3) for figure exemplifying gating the strategies described above.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Epigenetic gene silencing by heterochromatin primes fungal resistance

https://doi.org/10.1038/s41586-020-2706-x

Received: 3 October 2019

Accepted: 12 June 2020

Sito Torres-Garcia[1], Imtiyaz Yaseen[1], Manu Shukla[1], Pauline N. C. B. Audergon[1,2], Sharon A. White[1], Alison L. Pidoux[1] & Robin C. Allshire[1]✉

Heterochromatin that depends on histone H3 lysine 9 methylation (H3K9me) renders embedded genes transcriptionally silent[1–3]. In the fission yeast *Schizosaccharomyces pombe*, H3K9me heterochromatin can be transmitted through cell division provided the counteracting demethylase Epe1 is absent[4,5]. Heterochromatin heritability might allow wild-type cells under certain conditions to acquire epimutations, which could influence phenotype through unstable gene silencing rather than DNA change[6,7]. Here we show that heterochromatin-dependent epimutants resistant to caffeine arise in fission yeast grown with threshold levels of caffeine. Isolates with unstable resistance have distinct heterochromatin islands with reduced expression of embedded genes, including some whose mutation confers caffeine resistance. Forced heterochromatin formation at implicated loci confirms that resistance results from heterochromatin-mediated silencing. Our analyses reveal that epigenetic processes promote phenotypic plasticity, letting wild-type cells adapt to unfavourable environments without genetic alteration. In some isolates, subsequent or coincident gene-amplification events augment resistance. Caffeine affects two anti-silencing factors: Epe1 is downregulated, reducing its chromatin association, and a shortened isoform of Mst2 histone acetyltransferase is expressed. Thus, heterochromatin-dependent epimutation provides a bet-hedging strategy allowing cells to adapt transiently to insults while remaining genetically wild type. Isolates with unstable caffeine resistance show cross-resistance to antifungal agents, suggesting that related heterochromatin-dependent processes may contribute to resistance of plant and human fungal pathogens to such agents.

H3K9me-dependent heterochromatin can be copied by a read–write mechanism[4,5,8] and can arise stochastically at various loci, albeit only in the absence of key anti-silencing factors[9–13] or under specific growth conditions[14]. We reasoned that if heterochromatin can redistribute in wild-type *S. pombe* cells, it should be possible for epimutations to be generated, allowing adaptation to external insults. Unlike genetic mutants, we predicted that such epimutants would be unstable, resulting in gradual loss of resistance following growth without the insult. We chose to use caffeine as an insult because caffeine resistance is conferred by the deletion of genes with a variety of cellular roles[15], thus increasing the chance of obtaining epimutations. We also reasoned that unstable epimutants would occur more frequently at moderate caffeine concentrations that prevent most cells from growing (16 mM) than at the higher stringency (20 mM) used in screens for caffeine-resistant genetic mutants[15].

As secondary events might occur upon prolonged growth on caffeine, we froze an aliquot of each isolate upon formation of resistant colonies, as well as consecutive aliquots of each isolate after continued growth on caffeine (Fig. 1a). This ‘time series’ permitted the detection and separation of initiating events and potential subsequent changes. We therefore picked and froze colonies that grew after plating wild-type fission yeast (972 h^-) cells on 16 mM caffeine (+CAF). The resulting isolates were then successively propagated without caffeine (−CAF). Subsequently re-challenging these isolates with caffeine revealed that 23% lost caffeine resistance after 14 d of non-selective growth (UR, ‘unstable resistant’), 13% remained caffeine resistant (SR, ‘stable resistant’) and 64% did not display a clear phenotype (‘unclear’) (Fig. 1b and Extended Data Fig. 1a–c). Deleting *clr4*$^+$, which encodes the sole H3K9 methyltransferase of *S. pombe*[16,17], but not a control locus, from resistant isolates resulted in loss of caffeine resistance in unstable, but not stable, isolates (Fig. 1c and Extended Data Fig. 1d). Thus, caffeine resistance in unstable isolates is dependent on heterochromatin.

Whole-genome sequencing of a stable isolate, SR-1, uncovered a mutation in *pap1*$^+$ that was responsible for the caffeine-resistant phenotype[18] (Extended Data Fig. 1e). Chromatin immunoprecipitation sequencing (ChIP–seq) to detect dimethylation of H3K9 (H3K9me2) in SR-1 revealed no changes in heterochromatin distribution. Whole-genome sequencing of unstable isolates revealed no genetic changes in any sequence involved in either caffeine resistance or H3K9me2-mediated silencing, and 8 of 30 analysed unstable isolates had no detectable genetic

---

[1]Wellcome Centre for Cell Biology, Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK. [2]Present address: Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain. ✉e-mail: robin.allshire@ed.ac.uk

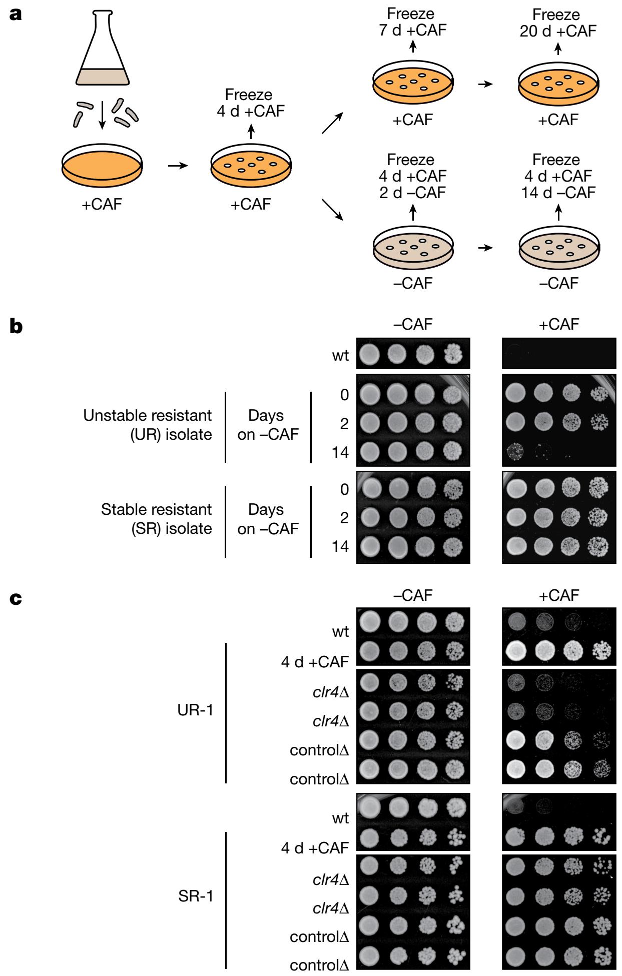


Fig. 1 | Identification of heterochromatin-dependent epimutants resistant to caffeine. a, Screening strategy. Wild-type (wt) *S. pombe* cells were plated on caffeine-containing (+CAF) medium, and caffeine-resistant colonies were picked and grown on +CAF medium for 4 d. Isolates from these colonies were then grown on +CAF medium for 7 or 20 d or on non-selective (−CAF) medium for 2 or 14 d. **b**, Growth of caffeine-resistant isolates after 2 or 14 d of non-selective growth. Isolates were serially diluted and spotted on −CAF and +CAF plates to assess the retention or loss of caffeine resistance, defining stable (SR) and unstable resistance (UR) status, respectively. **c**, Caffeine resistance in UR isolates depends on the Clr4 H3K9 methyltransferase. *clr4*$^+$ or an unlinked intergenic region was deleted (denoted *clr4Δ* and controlΔ, respectively) in unstable (UR-1) and stable (SR-1) caffeine-resistant isolates. Experiments in **b** and **c** were independently repeated at least twice with similar results.

changes from wild type (Extended Data Fig. 2a–e and Supplementary Table 1).

H3K9me2 ChIP–seq of unstable isolates, however, revealed altered heterochromatin distributions. Isolate UR-1 showed a new H3K9me2 island over the *hba1* locus, whereas UR-2 to UR-6 had H3K9me2 islands over *ncRNA.394*, *ppr4*, *grt1*, *fio1* and *mbx2*, respectively (Fig. 2 and Supplementary Table 1). Deletion of *hba1*$^+$ confers caffeine resistance[19], suggesting that caffeine-induced heterochromatin islands may drive resistance by silencing underlying genes. Accordingly, reverse transcription quantitative PCR (RT–qPCR) analysis revealed reduced expression of genes embedded in the observed *hba1* heterochromatin island (Extended Data Fig. 2f).

The *ncRNA.394*, *ppr4*, *grt1*, *fio1* and *mbx2* loci have not previously been implicated in caffeine resistance. Notably, however, 24 of 30 unstable isolates exhibited a heterochromatin island over the *ncRNA.394* locus (Extended Data Fig. 3a, b and Supplementary Table 1) and reduced transcript levels of embedded genes (Extended Data Figs. 2f, 3c), suggesting that transcriptional silencing within these loci mediates caffeine resistance.

ncRNA.394 was previously identified as a heterochromatin island that gains H3K9me2 in the absence of the counteracting Epe1 demethylase[9,20]. We detected no H3K9me2 over *ncRNA.394* in untreated wild-type cells (Fig. 2b and Extended Data Fig. 3a, b). Deletion of *ncRNA.394*$^+$ did not result in caffeine resistance (Extended Data Fig. 3d). Prolonged growth without caffeine of cells with the *ncRNA.394* heterochromatin island resulted in loss of H3K9me2 across this region, whereas their growth with caffeine led to the extension of the H3K9me2 domain over the genes *SPBC17G9.13c*$^+$ and *SPBC17G9.12c*$^+$ (Extended Data Fig. 3e). Deletion of *SPBC17G9.12c*$^+$ or *eno101*$^+$ did not result in caffeine resistance (Extended Data Fig. 3d). *SPBC17G9.13c*$^+$ is essential for viability, precluding testing its deletion for resistance.

To test whether heterochromatin formation over these loci alone results in caffeine resistance, we inserted *tetO* binding sites at *hba1*, *ncRNA.394* and *mbx2* to force the synthetic assembly of heterochromatin at these loci upon recruitment of the TetR–Clr4* fusion protein[4,5] (a tetracycline releasable version of Clr4 methyltransferase lacking the chromodomain). Combining *tetO* with TetR–Clr4* and growing cells without anhydrotetracycline (−AHT) induced TetR–Clr4* tethering to *tetO* and resulted in the formation of new H3K9me2 domains and their growth on caffeine (Fig. 3 and Extended Data Fig. 4a–d). Thus, heterochromatin-mediated silencing over *hba1*, *ncRNA.394* or *mbx2* results in caffeine resistance.

Notably, strains with forced synthetic heterochromatin at either *hba1* or *ncRNA.394* displayed resistance to the widely-used antifungals clotrimazole, tebuconazole and fluconazole (Fig. 3 and Extended Data Fig. 4e). Unstable caffeine-resistant isolates with heterochromatin islands over *hba1* (UR-1) or *ncRNA.394* (UR-2) also showed resistance to antifungals and produced small interfering RNAs (siRNAs) homologous to surrounding genes (Extended Data Fig. 5a–c). Because heterochromatin formation can involve the RNA interference (RNAi) pathway, we deleted RNAi components (*dcr1Δ* or *ago1Δ*) from UR-2 cells and found that their caffeine resistance was abolished (Extended Data Fig. 5d). Thus, RNAi also contributes to unstable caffeine resistance.

Tethering TetR–Clr4* near *SPBC17G9.13c*$^+$, upstream of *ncRNA.394*, resulted in caffeine resistance (Fig. 3c), suggesting that reduced expression of *SPBC17G9.13c*$^+$ (which we named *cup1*$^+$, for caffeine unstable phenotype 1) might mediate this resistance. We therefore created strains with manipulations that increased the degradation of *cup1*$^+$ mRNA (*LocusPX:cup1–3xDSR*) or attenuated its transcription (*cup1–TT*) (Methods). Both approaches resulted in less abundant *cup1*$^+$ transcripts and caffeine resistance (Extended Data Fig. 6a, b). Cup1 contains a LYR domain often found in mitochondrial proteins[21], and a Cup1–GFP fusion showed mitochondrial localization (Extended Data Fig. 6c). Mutation of the LYR domain led to caffeine resistance (Extended Data Fig. 6d). Thus, Cup1 (SPBC17G9.13c) is a mitochondrial protein whose mutation or reduced expression renders cells caffeine resistant. We conclude that silencing of wild-type *cup1*$^+$ due to the formation of a heterochromatin island mediates caffeine resistance in unstable isolates.

Besides the *ncRNA.394–cup1* heterochromatin island, analysis of ChIP–seq input DNA indicated that many independent isolates with unstable caffeine resistance also carried increased copies of a region of chromosome III (Extended Data Fig. 7a). The minimal region of overlap in 11 of 12 isolates contained *cds1*$^+$, whose overexpression confers caffeine resistance[22]. To determine whether *cds1*$^+$ became amplified before or after the formation of the *ncRNA.394–cup1* heterochromatin island, we analysed UR-2 samples frozen at different time points. We detected the *ncRNA.394–cup1* H3K9me2 island in the initial caffeine-resistant isolate (at 4 d +CAF), whereas amplification of the *cds1* locus arose later (at 7 d +CAF) (Extended Data Fig. 7b). Thus, the development of caffeine resistance appears to be a multistep process in which combinatorial events facilitate adaption to the insult.

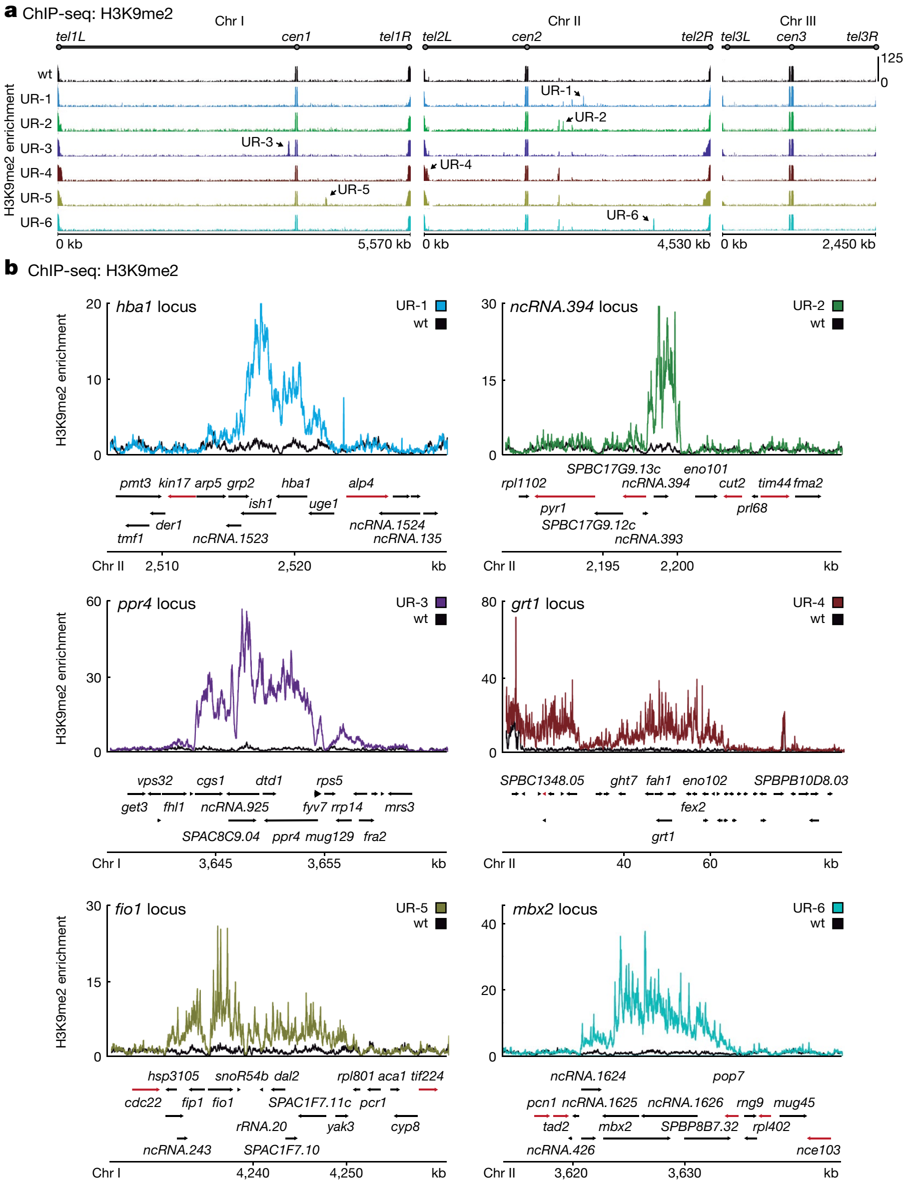


Fig. 2 | Ectopic islands of heterochromatin are detected in unstable (UR) caffeine-resistant isolates. a, **b**, Genome-wide (**a**) and locus-specific (**b**) H3K9me2 ChIP–seq enrichment in wild-type (wt) cells and unstable resistant (UR) isolates. Data are represented as relative fold enrichment over input. Sequencing was performed once, and results were confirmed by quantitative ChIP–qPCR (qChIP). Red arrows in **b** indicate essential genes.

In agreement with this hypothesis, deleting *clr4*$^+$ from the initial UR-2 isolate (4 d +CAF) resulted in loss of caffeine resistance in all transformants (6 of 6). However, only half (3 of 6; transformants 1, 4 and 5) lost caffeine resistance upon deletion of *clr4*$^+$ from the later UR-2 isolate (7 d +CAF) with *cds1* locus amplification. Transformants that retained resistance after *clr4*$^+$ was removed (3 of 6; transformants 2, 3 and 6) had a higher *cds1*$^+$ copy number than either *clr4*Δ transformants that lost resistance or wild-type cells (Extended Data Fig. 7c). We conclude that once *cds1* locus amplification occurs, heterochromatin is no longer required for caffeine resistance. In UR-2 the new *ncRNA.394–cup1* heterochromatin island arose before amplification of *cds1*$^+$, but it is likely that these events are stochastic and do not occur in a fixed order. Notably, both adaptations—island formation and locus amplification—were unstable and were lost following growth without caffeine (Extended Data Fig. 7d).

The instability of the amplified region suggests that the amplification resulted from excision and the formation of extrachromosomal circular DNA (eccDNA), structures that are prone to rapid accumulation and loss[23–26]. Copy number variation (CNV) plots revealed repetitive elements at the junctions of putative eccDNA (*5S rRNA.24–5S rRNA.26* for UR-2 only at 7 d +CAF and *LTR3–LTR27* for UR-4 at 4 d +CAF). PCR specific for putative circle junctions and Southern analysis confirmed the presence of eccDNA derived from chromosome III (Extended Data Fig. 8). Therefore, repeat-mediated generation of eccDNA is a potential alternative, or supplementary, mechanism for the evolution of resistance to caffeine, and perhaps other insults, in fission yeast. Accumulation of additional changes may allow further adaption to insults through other pathways or by bolstering silencing at particular loci[27].

To investigate the dynamics of heterochromatin island formation in response to caffeine, we exposed wild-type cells to low (7 mM) or medium (14 mM) doses of caffeine. Cells in low or medium caffeine doubled approximately eight or three times, respectively, in 18 h (Extended Data Fig. 9a). We detected several H3K9me2 heterochromatin islands after exposure to low caffeine (Fig. 4a, top, and Extended Data Fig. 9b, c). These represented a subgroup of the domains known to accumulate

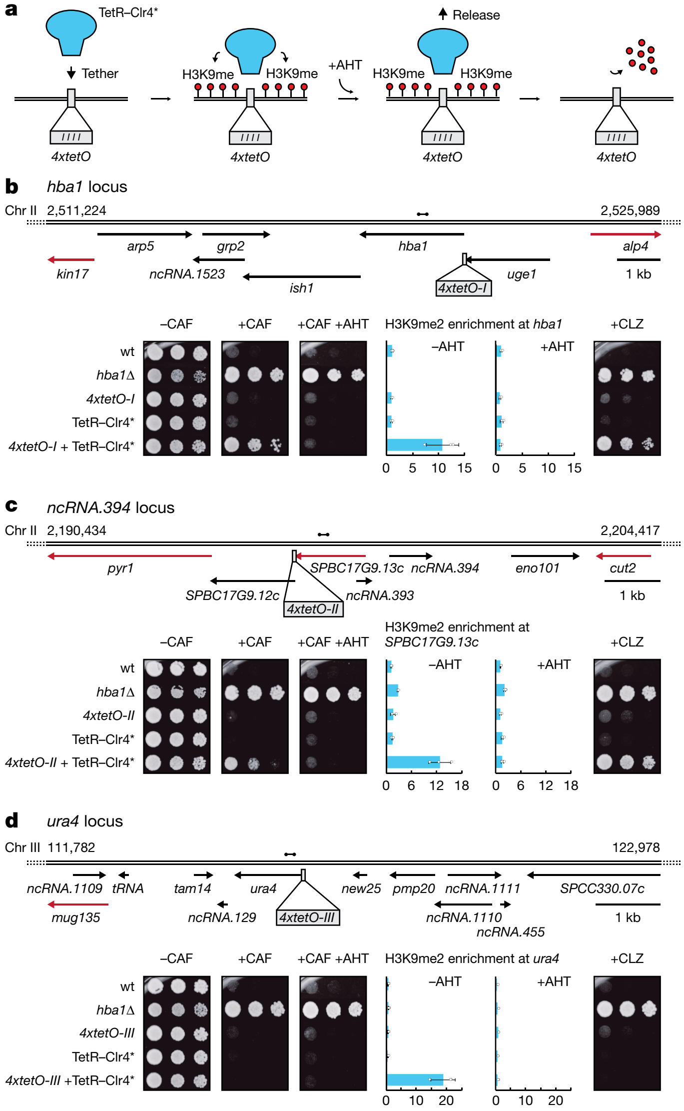


Fig. 3 | Forced synthetic heterochromatin at the identified loci is sufficient to drive caffeine resistance in wild-type cells. a, TetR–Clr4* mediates H3K9me deposition at *4xtetO* binding sites. Addition of anhydrotetracycline (+AHT) releases TetR–Clr4* from *4xtetO* sites, resulting in removal of H3K9me. **b**–**d**, Wild-type cells containing *4xtetO* binding sites at *hba1* or *ncRNA.394* (or *ura4* as control) and expressing TetR–Clr4* were assessed for caffeine (+CAF) or clotrimazole (+CLZ) resistance in the absence or presence of AHT. qChIP data for H3K9me2 levels at *hba1* (**b**), *SPBC17G9.13c* (near *ncRNA.394*; **c**) and *ura4* (**d**). Data are mean ± s.d. from three biological replicates. Dumbbells indicate primer pairs used. Red arrows indicate essential genes. *hba1Δ* denotes deletion of *hba1*$^+$.

H3K9me2 in the absence of Epe1[9,10,12], including *ncRNA.394*–*cup1*, but they did not overlap with H3K9me2-heterochromatin domains that accumulate in the absence of nuclear exosome function[13] or at 18 °C[14]. Notably, after medium caffeine treatment, ectopic heterochromatin was restricted to *ncRNA.394*–*cup1*, and H3K9me2 levels at this locus were approximately fourfold greater after exposure to medium compared to low caffeine (Fig. 4a and Extended Data Fig. 9d). Thus, exposure to the near-lethal 14 mM dosage of caffeine allows wild-type cells to develop resistance rapidly by forming heterochromatin at a locus (*ncRNA.394*–*cup1*) that confers resistance when silenced.

To determine whether other insults also induce the formation of heterochromatin islands, we exposed wild-type cells to oxidative stress (1 mM hydrogen peroxide). We detected heterochromatin islands at

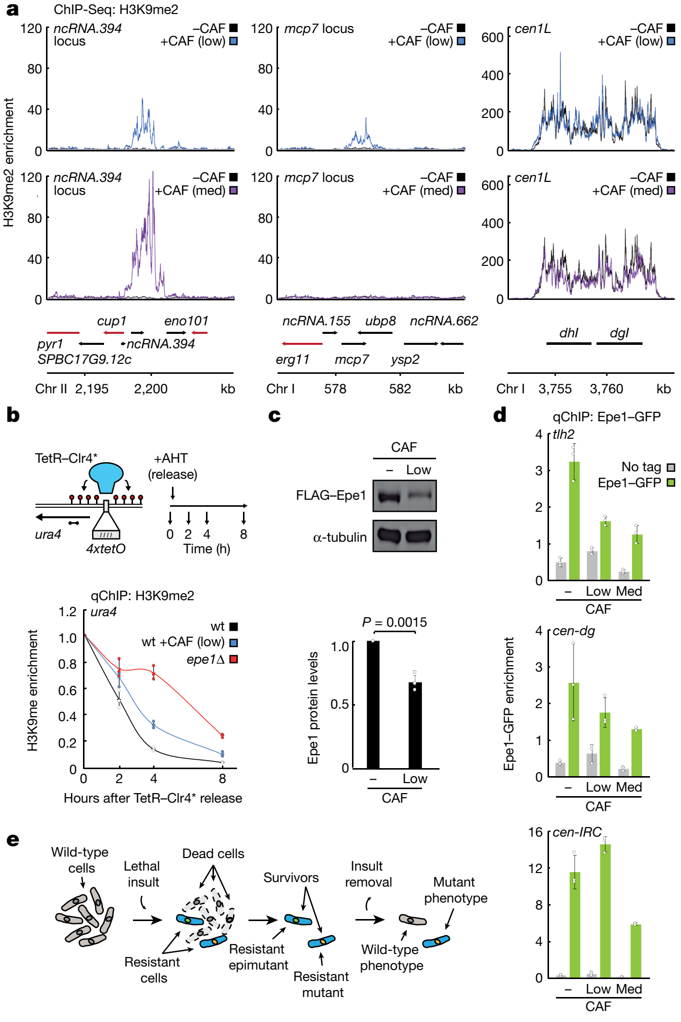


Fig. 4 | Dynamic heterochromatin redistribution following short exposure to caffeine in wild-type cells. a, H3K9me2 ChIP–seq enrichment at *ncRNA.394*–*cup1* and *mcp7* loci (or at pericentromeric *dgl*/*dhl* repeats of chromosome I as control) in wt cells following 18-h exposure to low (7 mM, top) or medium (14 mM, bottom) concentrations of caffeine. Data are represented as relative fold enrichment over input. Red arrows indicate essential genes. **b**, Effect of caffeine treatment on retention of synthetic heterochromatin upon release of tethered Clr4 methyltransferase. qChIP data for H3K9me2 levels (normalized to spike-in control) on *4xtetO-ura4*$^+$ before and after TetR–Clr4* release in wild-type cells treated with low caffeine, untreated control or *epe1Δ* cells (positive control). Dumbbells indicate primer pairs used. Data are mean ± s.d. from three biological replicates. **c**, Top, western analysis of 3×FLAG–Epe1 (endogenous gene tagged) levels before and after treatment with low caffeine. Loading control, α-tubulin. Gel source data shown in Supplementary Fig. 1a. Bottom, quantification of 3×FLAG–Epe1 protein levels normalized to α-tubulin. Data are mean ± s.d. of four biological replicates. *P* value, two-tailed Student's *t*-test. **d**, Effect of caffeine treatment on association of Epe1 with chromatin. qChIP analysis of Epe1–GFP levels at sub-telomeric *tlh2* locus and centromere 1 (*dg* repeats: cen-*dg*; outer boundary: cen-IRC) in wild-type cells treated with no, low or medium caffeine. Epe1–GFP levels were normalized to spike-in control. Data are mean ± s.d. from three biological replicates. **e**, Model. Resistant isolates arise following exposure to a lethal insult. Resistance could be mediated by permanent, DNA-based changes (resistant mutants) or reversible, heterochromatin-based epimutations (resistant epimutants). Upon removal of the insult, resistant epimutants can revert to wild type (sensitive phenotype) by disassembling ectopic heterochromatin islands, whereas resistant mutants continue displaying the mutant phenotype because of the genetic nature of DNA mutations.

locations similar to those observed in low caffeine, although H3K9me2 levels were lower (Extended Data Fig. 9b, c, e).

The heterochromatin profile of wild-type cells treated with low caffeine resembles that of untreated cells lacking Epe1 (*epe1*Δ) (Extended Data Fig. 9c). We hypothesized that caffeine might negatively regulate Epe1, thereby allowing adaptive ectopic heterochromatin islands to form in wild-type cells. TetR–Clr4*-mediated synthetic heterochromatin can be transmitted through cell division upon release of TetR–Clr4* from *tetO* sites only in cells lacking Epe1[4,5]. To further test whether caffeine imparts a phenotype similar to that of *epe1*Δ cells, we treated wild-type cells with low caffeine and released TetR–Clr4* from *tetO* sites inserted at *ura4*$^+$ (Fig. 4b). As in *epe1*Δ cells, caffeine treatment enabled heterochromatin to be retained at the tethering site for longer than in untreated cells. *epe1*$^+$ RNA levels were not substantially altered by caffeine, suggesting that this effect involves post-transcriptional regulation (Extended Data Fig. 9f). In cells expressing Epe1 fused to a 3×FLAG epitope tag, or GFP, from the endogenous *epe1* locus, exposure to caffeine resulted in 33% lower levels of 3×FLAG–Epe1 and reduced association of Epe1–GFP with various heterochromatin locations (Fig. 4c, d). These data suggest that downregulation of the putative H3K9 demethylase Epe1 has a crucial role in the response to external insults by allowing the formation of adaptive ectopic H3K9me-heterochromatin islands that, in turn, reduce the expression of underlying genes to confer resistance. Consistent with this scenario, *epe1*Δ cells formed more, and *clr4*Δ cells fewer, caffeine-resistant colonies than wild-type cells (Extended Data Fig. 9g).

Although caffeine reduces Epe1 protein levels, more H3K9me2 accumulated at heterochromatin islands in caffeine-treated wild-type cells than in untreated *epe1*Δ cells. (Extended Data Fig. 9c). Therefore, lower Epe1 levels alone cannot account for the high H3K9me2 observed at islands after caffeine treatment. The Mst2 histone acetyltransferase acts synergistically with Epe1 to prevent the formation of heterochromatin islands[10]. Notably, caffeine exposure caused wild-type cells to produce a shorter Mst2 protein (52 kDa versus 62 kDa) (Extended Data Fig. 10a). RNA sequencing results suggested that this shorter isoform arises through the use of an alternative transcriptional start site in cells exposed to caffeine, similar to effects observed with other stresses[28] (Extended Data Fig. 10b). We suggest that this caffeine-induced shorter isoform, which lacks the MYST–zinc finger domain normally found in Mst2[29], may be inactive and unable to prevent heterochromatin island formation. Thus, caffeine treatment of wild-type cells, both by lowering Epe1 levels and probably by disabling Mst2, allows greater accumulation of H3K9me2 at islands than is seen in *epe1*Δ cells. These findings reveal an adaptive epigenetic response to external insults that stimulates phenotypic plasticity, and suggest that stress-response pathways may regulate heterochromatin modulation activities, thereby ensuring cell survival in fluctuating environmental conditions (Fig. 4e).

Epimutations dependent on 5-methycytosine (5-meC) DNA methylation frequently arise in plants and are propagated by maintenance methyltransferases[30,31]. RNAi-mediated epimutations occur in the fungus *Mucor circinelloides*[32], but whether this process is dependent on DNA methylation or heterochromatin is unknown. As fission yeast lacks 5-meC DNA methylation[33,34], this epigenetic mark cannot be responsible for the epimutations described here. Instead our analyses indicate that these adaptive epimutations are transmitted in wild-type cells by the Clr4–H3K9me read–write mechanism[4,5,8].

Our results raise the question of why epimutants have not been detected in previous mutant screens performed in fission yeast. Stringent phenotypic screens mean that strong mutants are investigated further and eccentric mutants discarded. Here, however, we selected for weak mutants by applying sublethal doses of drug at the threshold of growth prevention. Selection was time limited to maximize identification of isolates showing unstable phenotypes before development of genetic alterations.

Fungal infections are on the rise, especially in immunocompromised humans, yet few effective antifungal agents exist, and resistance is rendering these increasingly ineffective[35]. Widespread use of related azole compounds to control fungus-mediated crop deterioration may leave residual antifungals in the soil, possibly allowing unwitting selection of resistant epimutants in fungi and ultimately driving increasing cases of azole-resistant aspergillosis and cryptococcosis in humans. Monitoring resistance in clinical isolates involves identifying mutations by genome sequencing[36], but this would miss resistance due to epimutations such as those described here, leading to inaccurate diagnoses. Re-engineering existing so-called 'epigenetic drugs'—compounds that inhibit histone-modifying enzymes—or searching for novel agents of this type may identify molecules that specifically block the formation of fungal, but not host, heterochromatin, reducing the emergence of antifungal resistance in clinical and agricultural settings.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2706-x.

1. Bannister, A. J. et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
2. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
3. Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol*. **19**, 229–244 (2018).
4. Audergon, P. N. C. B. et al. Epigenetics. Restricted epigenetic inheritance of H3K9 methylation. *Science* **348**, 132–135 (2015).
5. Ragunathan, K., Jih, G. & Moazed, D. Epigenetics. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science* **348**, 1258699 (2015).
6. Jeggo, P. A. & Holliday, R. Azacytidine-induced reactivation of a DNA repair gene in Chinese hamster ovary cells. *Mol. Cell. Biol*. **6**, 2944–2949 (1986).
7. Oey, H. & Whitelaw, E. On the meaning of the word 'epimutation'. *Trends Genet*. **30**, 519–520 (2014).
8. Zhang, K., Mosch, K., Fischle, W. & Grewal, S. I. S. Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat. Struct. Mol. Biol*. **15**, 381–388 (2008).
9. Zofall, M. et al. RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. *Science* **335**, 96–100 (2012).
10. Wang, J., Reddy, B. D. & Jia, S. Rapid epigenetic adaptation to uncontrolled heterochromatin spreading. *eLife* **4**, e06179 (2015).
11. Parsa, J.-Y., Boudoukha, S., Burke, J., Homer, C. & Madhani, H. D. Polymerase pausing induced by sequence-specific RNA-binding protein drives heterochromatin assembly. *Genes Dev*. **32**, 953–964 (2018).
12. Sorida, M. et al. Regulation of ectopic heterochromatin-mediated epigenetic diversification by the JmjC family protein Epe1. *PLoS Genet*. **15**, e1008129 (2019).
13. Yamanaka, S. et al. RNAi triggered by specialized machinery silences developmental genes and retrotransposons. *Nature* **493**, 557–560 (2013).
14. Gallagher, P. S. et al. Iron homeostasis regulates facultative heterochromatin assembly in adaptive genome control. *Nat. Struct. Mol. Biol*. **25**, 372–383 (2018).
15. Calvo, I. A. et al. Genome-wide screen of genes required for caffeine tolerance in fission yeast. *PLoS ONE* **4**, e6619 (2009).
16. Ivanova, A. V., Bonaduce, M. J., Ivanov, S. V. & Klar, A. J. The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast. *Nat. Genet*. **19**, 192–195 (1998).
17. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113 (2001).
18. Kudo, N., Taoka, H., Toda, T., Yoshida, M. & Horinouchi, S. A novel nuclear export signal sensitive to oxidative stress in the fission yeast transcription factor Pap1. *J. Biol. Chem*. **274**, 15151–15158 (1999).
19. Castillo, E. A., Vivancos, A. P., Jones, N., Ayté, J. & Hidalgo, E. *Schizosaccharomyces pombe* cells lacking the Ran-binding protein Hba1 show a multidrug resistance phenotype due to constitutive nuclear accumulation of Pap1. *J. Biol. Chem*. **278**, 40565–40572 (2003).
20. Zofall, M., Smith, D. R., Mizuguchi, T., Dhakshnamoorthy, J. & Grewal, S. I. S. Taz1-Shelterin promotes facultative heterochromatin assembly at chromosome-internal sites containing late replication origins. *Mol. Cell* **62**, 862–874 (2016).
21. Angerer, H. Eukaryotic LYR proteins interact with mitochondrial protein complexes. *Biology (Basel)* **4**, 133–150 (2015).
22. Wang, S. W., Norbury, C., Harris, A. L. & Toda, T. Caffeine can override the S-M checkpoint in fission yeast. *J. Cell Sci*. **112**, 927–937 (1999).
23. Libuda, D. E. & Winston, F. Amplification of histone genes by circular chromosome formation in *Saccharomyces cerevisiae*. *Nature* **443**, 1003–1007 (2006).

24. Møller, H. D., Parsons, L., Jørgensen, T. S., Botstein, D. & Regenberg, B. Extrachromosomal circular DNA is common in yeast. *Proc. Natl Acad. Sci. USA* **112**, E3114–E3122 (2015).
25. Hull, R. M. et al. Transcription-induced formation of extrachromosomal DNA during yeast ageing. *PLoS Biol*. **17**, e3000471 (2019).
26. Wu, S. et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 (2019).
27. Stajic, D., Perfeito, L. & Jansen, L. E. T. Epigenetic gene silencing alters the mechanisms and rate of evolutionary adaptation. *Nat. Ecol. Evol*. **3**, 491–498 (2019).
28. Thodberg, M. et al. Comprehensive profiling of the fission yeast transcription start site activity during stress and media response. *Nucleic Acids Res*. **47**, 1671–1691 (2019).
29. Yan, Y., Barlev, N. A., Haley, R. H., Berger, S. L. & Marmorstein, R. Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases. *Mol. Cell* **6**, 1195–1205 (2000).
30. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**, 157–161 (1999).
31. Heard, E. & Martienssen, R. A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157**, 95–109 (2014).
32. Calo, S. et al. Antifungal drug resistance evoked via RNAi-dependent epimutations. *Nature* **513**, 555–558 (2014).
33. Antequera, F., Tamame, M., Villanueva, J. R. & Santos, T. DNA methylation in the fungi. *J. Biol. Chem*. **259**, 8033–8036 (1984).
34. Wilkinson, C. R., Bartlett, R., Nurse, P. & Bird, A. P. The fission yeast gene *pmt1*$^+$ encodes a DNA methyltransferase homologue. *Nucleic Acids Res*. **23**, 203–210 (1995).
35. Fisher, M. C., Hawkins, N. J., Sanglard, D. & Gurr, S. J. Worldwide emergence of resistance to antifungal drugs challenges human health and food security. *Science* **360**, 739–742 (2018).
36. Stone, N. R. et al. Dynamic ploidy changes drive fluconazole resistance in human cryptococcal meningitis. *J. Clin. Invest*. **129**, 999–1014 (2019).

Methods

Yeast strains and manipulations

Standard methods were used for fission yeast growth, genetics and manipulation[37]. *S. pombe* strains used in this study are described in Supplementary Table 2. Oligonucleotide sequences are listed in Supplementary Table 3. For pDUAL–adh21–TetR–2xFLAG–Clr4–CDΔ (abbreviated as TetR–Clr4*), the *nmt81* promoter of pDUAL–nmt81–TetR–2xFLAG–Clr4–CDΔ[4] was replaced by the *adh21* promoter (pRAD21, gift from Y. Watanabe). The *Not*I-digested plasmid was integrated at *leu1*$^+$.

To reduce the expression of *SPBC17G9.13c*$^+$/*cup1*$^+$, we used two independent strategies. First, we expressed an additional copy of *cup1*$^+$ with three DSR (determinant of selective removal) nuclear exosome RNA degradation motifs[38,39] fused to its 3′ untranslated region from an intergenic locus (*LocusPX:cup1-3xDSR*). Following insertion of *cup1–3xDSR* at *LocusPX*, endogenous *cup1*$^+$ was deleted and cells expressing only *cup1–3xDSR* were analysed. Second, the 144-bp transcriptional terminator site from *ura4*$^+$ was inserted in place of part of the putative *cup1*$^+$ promoter (*cup1–TT*) and cells were analysed.

pap1–N424STOP, *clr5–Q264STOP meu27–S100Y*, *LocusPX:cup1–3xDSR*, *cup1–TT*, *cup1–L73G*, *cup1–F99G*, *cup1–GFP*, *3xFLAG–epe1* and strains carrying *4xtetO* insertions were constructed by CRISPR–Cas9-mediated genome editing using the *SpEDIT* system (Allshire Laboratory; available on request) with oligonucleotides listed in Supplementary Table 3. The mitochondrial protein Arg11[40], Epe1 and Mst2 were C-terminally tagged with mCherry (Arg11), GFP (Epe1) or 13×Myc (Mst2) using the Bähler tagging method[41]. Most gene deletions, including *ago1*Δ, *clr4*Δ and *dcr1*Δ, were made by the standard Bähler deletion method[41]; *epe1*Δ was constructed using the *SpEDIT* genome editing system.

Yeast extract plus supplements (YES) was used to grow all cultures. Sixteen mM caffeine (Sigma, C0750) was added to medium for caffeine-resistance screens and serial dilution assays. To screen for unstable caffeine-resistant isolates, caffeine-resistant colonies that formed 7 d after plating of wild-type cells on 16 mM caffeine YES (+CAF) plates were picked and patched to +CAF plates. After 4 d of growth, isolates were frozen (4 d +CAF). Four d +CAF isolates were re-patched and grown for 3 d on +CAF plates and then frozen (7 d +CAF). Subsequently, 7 d +CAF isolates were re-patched every 3 d on +CAF plates up to 20 d of total growth on +CAF plates and then frozen (20 d +CAF).

Clotrimazole (0.29 µM) (Sigma, C6019) was added to medium for clotrimazole resistance serial dilution assays. Tebuconazole (1.6 µM) (Sigma, 32013) was added to medium for tebuconazole resistance serial dilution assays. Fluconazole (0.6 mM) (Sigma, PHR1160) was added to medium for fluconazole resistance serial dilution assays.

Seven or 14 mM caffeine (Sigma, C0750) or 1 mM hydrogen peroxide (Sigma, H1009) were added to medium for 18 h for drug treatment experiments. To release TetR–Clr4*, 10 µM anhydrotetracycline (AHT) was added to the medium.

Serial dilution assays

Equal amounts of starting cells were serially diluted fivefold and then spotted onto appropriate medium. Cells were grown at 30–32 °C for 3–5 d and then photographed.

Chromatin immunoprecipitation (ChIP)

ChIP experiments were performed as previously described[42] using anti-H3K9me2 (5.1.1, gift from T. Urano) or anti-GFP (Invitrogen, A11122). Immunoprecipitated DNA was recovered with Chelex-100 resin (Bio-Rad) for ChIP–qPCR (qChIP) experiments or with QIAquick PCR Purification Kit (Qiagen) for ChIP–seq experiments.

Quantitative ChIP–qPCR (qChIP)

qChIP data were analysed by real-time PCR using Lightcycler 480 SYBR Green (Roche) with oligonucleotides listed in Supplementary Table 3. All ChIP enrichments were calculated as % DNA immunoprecipitated at the locus of interest relative to the corresponding input samples and normalized to % DNA immunoprecipitated at the *act1*$^+$ locus. For spike-in qChIPs, an equal number (about 20%) of *Schizosaccharomyces octosporus* cells (H3K9me2 spike-in qChIP)[42] or Sgo1–GFP *Saccharomyces cerevisiae* cells (GFP spike-in qChIP)[43] (gift from A. Marston) were added to initial *S. pombe* pellets. Histograms represent data averaged over three biological replicates. Error bars represent standard deviations.

ChIP–seq library preparation and analysis

Illumina-compatible libraries were prepared as previously described[42] using NEXTflex-96 barcode adapters (Bioo Scientific) and Ampure XP beads (Beckman Coulter). Libraries were then pooled to allow multiplexing and sequenced on an Illumina HiSeq2000, NextSeq or MiniSeq system (150-cycle high-output kit) by 75-bp paired-end sequencing.

Approximately 6–10 million 75-bp paired-end reads were produced for each sample. Raw reads were then de-multiplexed and trimmed using Trimmomatic (v0.35)[44] to remove adaptor contamination and regions of poor sequencing quality. Trimmed reads were aligned to the *S. pombe* reference genome (*972h*$^-$, ASM294v2.20) using Bowtie2 (v2.3.3)[45]. Resulting bam files were processed using Samtools (v1.3.1)[46] and Picard Tools (v2.1.0) (http://broadinstitute.github.io/picard) for sorting, removing duplicates and indexing. Coverage bigwig files were generated by BamCoverage (deepTools v2.0), and IP/input ratios were calculated using BamCompare (deepTools v2.0)[47] in SES mode for normalization[48]. Peaks were called using MACS2[49] in PE mode and broad peak calling (broad-cutoff = 0.05). Region-specific H3K9me2 enrichment plots were generated using the Sushi R package (v1.22)[50]. Heat maps were generated using computeMatrix and plotHeatmap (deepTools v2.0)[47] with genomic coordinates indicated in Supplementary Table 4.

SNP and indel calling

SNPs and insertions/deletions (indels) were called as previously described[51]. Trimmed reads were mapped to the *S. pombe* reference genome (*972h*$^-$, ASM294v2.20) using Bowtie2 (v2.3.3)[45]. GATK[52,53] was used for base quality score recalibration. SNPs and indels were called with GATK HaplotypeCaller[52,53] and filtered using custom parameters. Functional effect of variants was determined using Variant Effect Predictor[54].

Copy number variation analysis

Copy number variation (CNV) was determined using CNVkit[55] in Whole-Genome Sequencing (-wgs) mode. Wild-type ChIP–seq input bam files were used as reference.

Extrachromosomal circular DNA diagnostic PCRs and Southern analysis

ChIP-input DNA samples were used as template for PCR with Taq polymerase (Roche, 4728858001) according to manufacturer's instructions. Two types of PCR were performed: control PCR for loci present on endogenous chromosome III (expected to be present in wild-type, UR-2 (7 d +CAF) and UR-4) and circle-specific PCRs specific for putative extrachromosomal circles predicted to be present in UR-2 (7 d +CAF) or UR-4. For wild-type and UR-2 (7 d +CAF): control primers were located on either on side of *5S rRNA.24* (primers A (forward), B (reverse); Supplementary Table 3) and *5S rRNA.26* (primers C, D); circle-specific primers were located on either side of a predicted junction between *5S rRNA.24* and *5S rRNA.26* (primers C and B). For wild-type and UR-4: control primers were located on one side of either *LTR3* (primers E, F) or *LTR27* (primers G, H); circle-specific primers were located on either side of a predicted junction between *LTR3* and *LTR27* (primers G and F). For some locations, more than one forward and/or reverse primer was used: for instance, forward primers C1, C2 with reverse primers D1, D2. PCR products were electrophoresed on 2% agarose gels containing ethidium bromide.

For Southern blot analysis, genomic DNA was prepared from wild-type, UR-2 (7 d +CAF) and UR-4 cultures grown in YES. In brief, cells were incubated with Zymolyase 100T (AMS Biotechnology) to digest the cell wall, pelleted, resuspended in TE and lysed with sodium dodecyl sulfate, and then potassium acetate was added and the lysates precipitated with isopropanol. After treatment with RNase A and proteinase K, phenol/chloroform and chloroform extractions were performed. DNA was precipitated in the presence of sodium acetate and ethanol, which was followed by centrifugation and washing of the pellet with 70% ethanol. After air drying, the pellet was resuspended in TE. Approximately 8 µg of DNA was digested with the following restriction enzymes: wild type and UR-2 (7 d +CAF): *Bsm*BI, *Eco*RV, *Nde*I; wild type and UR-4: *Eco*RI, *Bam*HI + *Xba*I. Digested DNA was subjected to electrophoresis in a 0.9% agarose gel containing ethidium bromide. Southern blotting was achieved by the alkali transfer method. In brief, the gel was depurinated with 0.3 M HCl for 10 min, washed with distilled water, and incubated twice for 15 min each in denaturing solution (0.5 M NaOH, 1.5 M NaCl). Overnight capillary transfer was used for transfer to Hybond XL membrane (Amersham), which was then washed with 50 mM Na_2HPO_4 pH7.2, followed by air drying. After drying at 80 °C for 2 h and UV crosslinking, membranes were prehybridized in Church buffer (0.5 M Na_2HPO_4 pH 7.2, 7% SDS, 1 mM EDTA, 1% BSA (Sigma, A0281) for 1 h at 65 °C. Probes were made using the High Prime kit (Roche, 11585592001) and α-^{32}P-dCTP (NEN), according to the manufacturer's instructions. Heat denatured probes in Church buffer were hybridized with relevant membranes at 65 °C overnight with rotation. Following washes with wash buffer (40 mM Na_2HPO_4 pH 7.2, 1 mM EDTA, 1% SDS), blots were exposed to XAR-5 film (Kodak) at −80 °C with an intensifying screen for several hours.

Cytology

Schizosaccharomyces pombe cultures were fixed before processing for immunofluorescence as described[42]. In brief, cells in YES culture were fixed with 3.7% formaldehyde (Sigma, F8775) for 30 min, followed by cell wall digestion with Zymolyase-100T (AMS Biotechnology) in PEMS buffer (100 mM PIPES pH 7, 1 mM EDTA, 1 mM $MgCl_2$, 1.2 M Sorbitol). After permeabilization with Triton-X100, cells were washed and blocked in PEMBAL (PEM containing 1% BSA, 0.1% sodium azide, 100 mM lysine hydrochloride). Rabbit anti-GFP (Invitrogen, A11122) was used in PEMBAL at 1:500 dilution, and Alexa-488-coupled chicken anti-rabbit secondary antibody (Invitrogen, A21441) at 1:1,000 dilution. Arg11–mCherry fluorescence survived fixation, and no antibodies were used for localization. Cells were stained with DAPI and mounted in Vectashield. Microscopy was performed with a Zeiss Imaging 2 microscope (Zeiss) using a 100×, 1.4-NA Plan-Apochromat objective, Prior filter wheel and illumination by HBO100 mercury bulb. Image acquisition with a Photometrics Prime sCMOS camera (Photometrics, https://www.photometrics.com) was controlled using Metamorph software (Version 7; Universal Imaging). Exposures were 3000 ms for FITC/Alexa-488 channel (Cup1–GFP/Alexa 488), 500 ms for TRITC channel (Arg11–mCherry) and 100 ms for DAPI. For display of images, maximum intensity was determined for, for example, Cup1–GFP staining in Cup1–GFP Arg11–mCherry strain (B4909), and this maximum was applied for scaling of all B4909 and B4912 (which expresses only Arg11–mCherry) images. FITC and TRITC channels were scaled in this way; DAPI images were autoscaled.

qRT–PCR analysis

Total RNA was extracted using the Monarch Total RNA Miniprep Kit (New England Biolabs) according to the manufacturer's instructions. Contaminating DNA was removed by treating with Turbo DNase (Invitrogen), and reverse transcription was performed using LunaScript RT Supermix Kit (New England Biolabs). Oligonucleotides used for qRT–PCR are listed in Supplementary Table 3. qRT–PCR histograms represent three biological replicates; error bars correspond to the standard deviation.

RNA sequencing library preparation and analysis

Total RNA was extracted using the Monarch Total RNA Miniprep Kit (New England Biolabs) according to the manufacturer's instructions. Contaminating DNA was removed by treating with Turbo DNase (Invitrogen). rRNA was removed using the Ribo-Zero Gold rRNA removal kit (Yeast) (Illumina) before library construction using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). Libraries were pooled and sequenced on an Illumina NextSeq platform by 75-bp paired-end sequencing. Adaptor-trimmed reads were aligned to the *S. pombe* reference genome (*972h*$^-$, ASM294v2.20) using STAR[56] (v2.2.1) and processed using Samtools[46] (v1.3.1). Coverage bigwig files were generated by BamCoverage (deepTools v2.0)[47].

Differential expression was analysed using the Bioconductor Rsamtools (v2.0.3), GenomicFeatures[57] (v1.36.4) and DESeq2[58] (v.1.24) R libraries. $\log_2$-transformed fold changes were shrunk using the apeglm method[59] and a MA plot was generated using R. Genes with an adjusted *P* value <0.01 are shown in red.

Small RNA sequencing

Fifty ml of log-phase cells were collected and processed using the mirVana miRNA Isolation kit (Invitrogen). Resulting sRNA was treated with TURBO DNase (Invitrogen) and used for library construction with the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs) according to manufacturer's instructions. Libraries were pooled and sequenced on an Illumina NextSeq platform by 50-bp single-end sequencing. Raw reads were then de-multiplexed and processed using Cutadapt (v1.17) to remove adaptor contamination and discard reads shorter than 19 nucleotides or longer than 25 nucleotides. Coverage plots were generated using SCRAM[60].

Protein extraction and western analysis

Protein samples were prepared as previously detailed[61]. Western detection was performed using anti-FLAG–HRP (Sigma, A8591), anti-Myc (Cell Signalling, 9B11), anti-α-tubulin[62] (gift from K. Gull), goat anti-mouse (Sigma, A4416), anti-Bip1[63], goat anti-rabbit (Sigma, A6154), anti-Cdc11 (gift from K. Sawin) and donkey anti-sheep (Abcam, ab6900). Gels were visualized using the ChemiDoc imaging system (Bio-Rad) and analysed with ImageJ.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequence data generated in this study have been submitted to GEO under accession number GSE138436. Source data are provided with this paper.

Code availability

The complete Workflow Description Language (WDL) pipeline script used for ChIP–seq and variation analyses is available at https://github.com/SitoTorres/Torres-Garcia-et-al.-2019.

37. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
38. Harigaya, Y. et al. Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* **442**, 45–50 (2006).
39. Watson, A. T. et al. Optimisation of the *Schizosaccharomyces pombe urg1* expression system. *PLoS ONE* **8**, e83800 (2013).
40. Delerue, T. et al. Loss of Msp1p in *Schizosaccharomyces pombe* induces a ROS-dependent nuclear mutator phenotype that affects mitochondrial fission genes. *FEBS Lett.* **590**, 3544–3558 (2016).
41. Bähler, J. et al. Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**, 943–951 (1998).

42. Tong, P. et al. Interspecies conservation of organisation and function between nonhomologous regional centromeres. *Nat. Commun*. **10**, 2343 (2019).
43. Nerusheva, O. O., Galander, S., Fernius, J., Kelly, D. & Marston, A. L. Tension-dependent removal of pericentromeric shugoshin is an indicator of sister chromosome biorientation. *Genes Dev*. **28**, 1291–1309 (2014).
44. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
46. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
47. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. **44** (W1), W160–W165 (2016).
48. Diaz, A., Park, K., Lim, D. A. & Song, J. S. Normalization, bias correction, and peak calling for ChIP-seq. *Stat. Appl. Genet. Mol. Biol*. **11**, 9 (2012).
49. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. **9**, R137 (2008).
50. Phanstiel, D. H., Boyle, A. P., Araya, C. L. & Snyder, M. P. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics* **30**, 2808–2810 (2014).
51. Jeffares, D. C. et al. The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nat. Genet*. **47**, 235–241 (2015).
52. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. **20**, 1297–1303 (2010).
53. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*. **43**, 11.10.1–11.10.33 (2013).
54. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol*. **17**, 122 (2016).
55. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol*. **12**, e1004873 (2016).
56. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
57. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLOS Comput. Biol*. **9**, e1003118 (2013).
58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**, 550 (2014).
59. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).
60. Fletcher, S. J., Boden, M., Mitter, N. & Carroll, B. J. SCRAM: a pipeline for fast index-free small RNA read alignment and visualization. *Bioinformatics* **34**, 2670–2672 (2018).
61. Braun, S. et al. The Cul4-Ddb1(Cdt)^2 ubiquitin ligase inhibits invasion of a boundary-associated antisilencing factor into heterochromatin. *Cell* **144**, 41–54 (2011).
62. Woods, A. et al. Definition of individual components within the cytoskeleton of *Trypanosoma brucei* by a library of monoclonal antibodies. *J. Cell Sci*. **93**, 491–500 (1989).
63. Pidoux, A. L. & Armstrong, J. The BiP protein and the endoplasmic reticulum of *Schizosaccharomyces pombe*: fate of the nuclear envelope during cell division. *J. Cell Sci*. **105**, 1115–1120 (1993).

Acknowledgements We thank L. Di Pompeo, A. Fellas and R. Yeboah for laboratory support; P. Tong, M. Lafos, R. Ard and S. Webb (Wellcome Centre for Cell Biology Bioinformatics Core) for sharing technical expertise; D. Kelly (Wellcome Centre Optical Instrumentation Laboratory) for microscopy and instrumentation support; members of the Allshire laboratory for valuable discussions; A. Bird, W. Bickmore and L. Massari for comments on the manuscript; T. Urano for the 5.1.1 (H3K9me) antibody; Y. Watanabe for the pRAD21 plasmid; K. Gull for the α-tubulin antibody; A. Marston for the *Sgo1–GFP S. cerevisiae* strain, K. Sawin for the Cdc11 antibody; and Edinburgh Genomics (NERC, R8/H10/56; MRC, MR/K001744/1; BBSRC, BB/J004243/1) and the Genetics Core, Edinburgh Clinical Research Facility, University of Edinburgh, for sequencing. S.T.-G. is supported by the Darwin Trust of Edinburgh. P.N.C.B.A. was supported by the Wellcome 4 Year PhD program in Cell Biology (093852). I.Y. is supported by an EMBO Long Term Fellowship (EMBO ALTF 130-2018). R.C.A. is a Wellcome Principal Research Fellow (095021, 200885); the Wellcome Centre for Cell Biology is supported by core funding from Wellcome (203149).

Author contributions S.T.-G., P.N.C.B.A. and R.C.A. conceived the project. S.T.-G. and P.N.C.B.A. performed preliminary studies. S.T.-G. performed experiments and bioinformatics. M.S. designed *cup1-3xDSR* experiments and contributed to ChIP–seq and qChIP experiments. A.L.P. performed cytology, *cup1-TT* and eccDNA experiments. I.Y. constructed strains expressing epitope-tagged Epe1 and Mst2 and performed western analysis. S.A.W. generated strains expressing Cup1–GFP and mutant Cup1 and contributed to Epe1 and Mst2 experiments. S.T.-G., A.L.P. and R.C.A. wrote the manuscript.

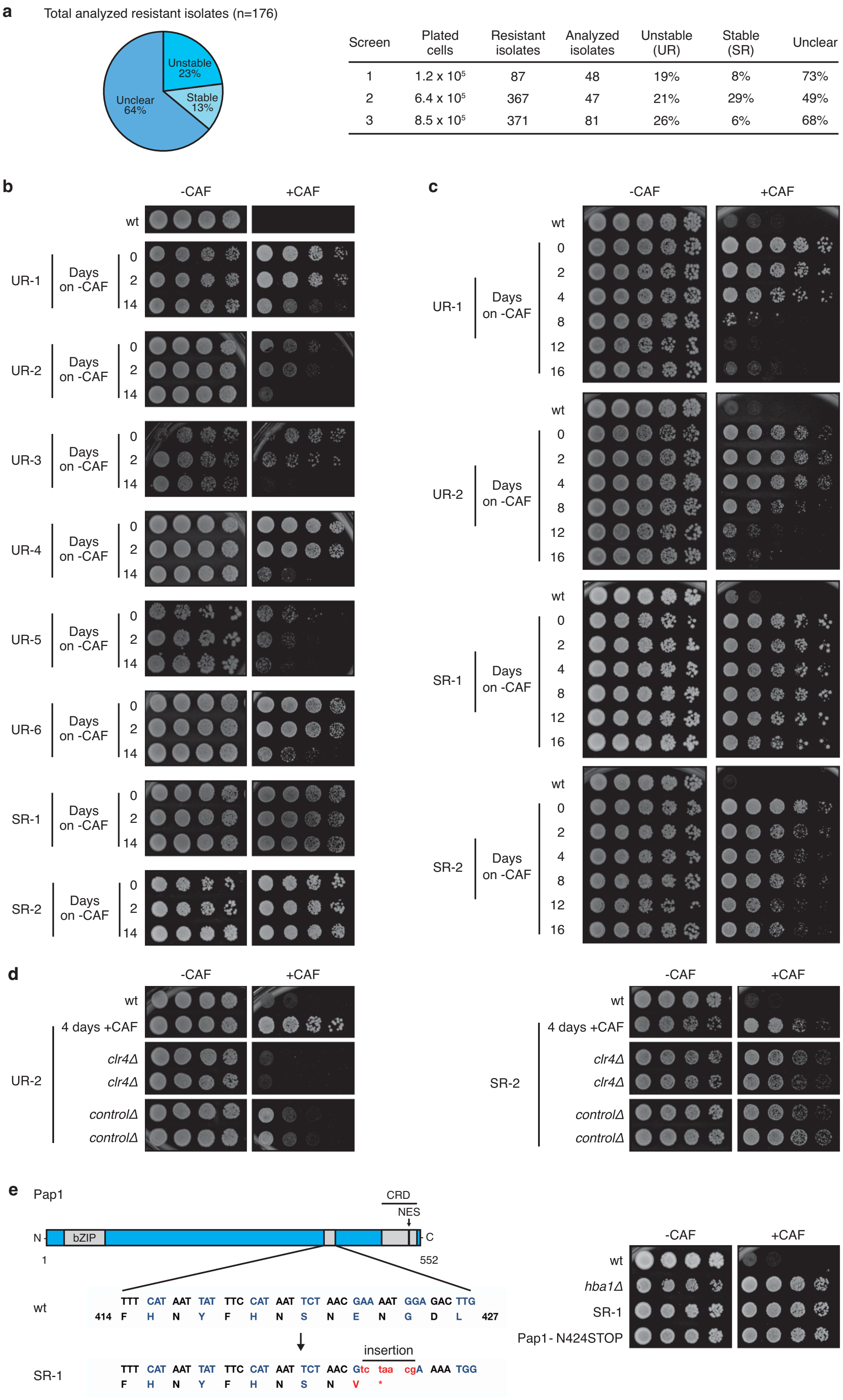
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2706-x.

Correspondence and requests for materials should be addressed to R.C.A.

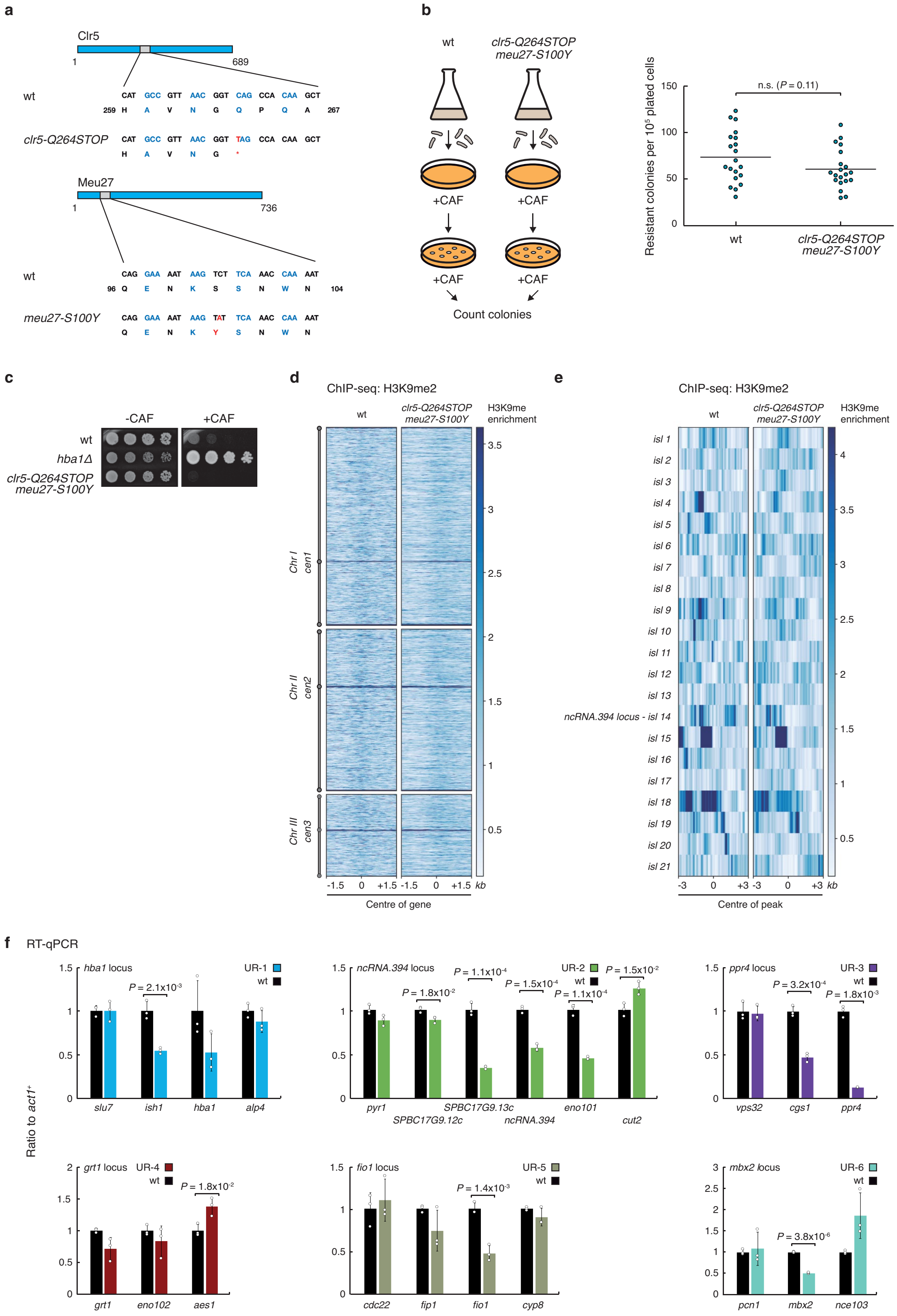
Peer review information *Nature* thanks Songtao Jia and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



Screen	Plated cells	Resistant isolates	Analyzed isolates	Unstable (UR)	Stable (SR)	Unclear
1	1.2×10^5	87	48	19%	8%	73%
2	6.4×10^5	367	47	21%	29%	49%
3	8.5×10^5	371	81	26%	6%	68%

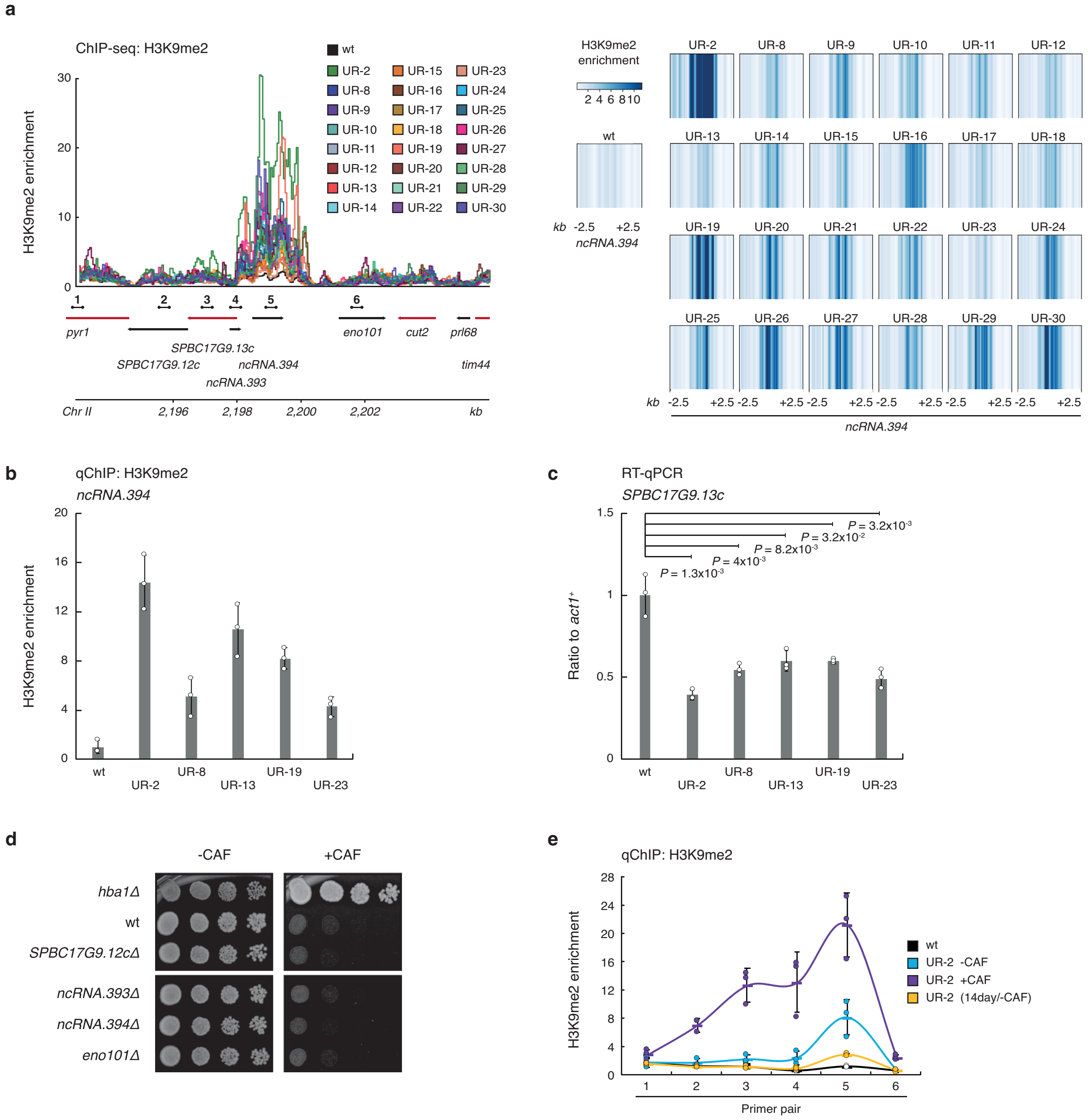
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Identification of heterochromatin-dependent epimutants resistant to caffeine. a, Frequencies of unstable (UR) and stable (SR) caffeine-resistant isolates obtained from three independent screens. 64% of isolates did not display a clear phenotype (unclear). **b**, Unstable (UR) and stable (SR) caffeine-resistant isolates were identified using this screening strategy. After growth on non-selective media for 14 d, caffeine resistance is lost in UR isolates but not in SR isolates. **c**, Caffeine resistance is lost progressively in unstable (UR) isolates but maintained in stable (SR) isolates. **d**, Caffeine resistance in UR isolates depends on the Clr4 H3K9 methyltransferase. *clr4*$^+$ (*clr4Δ*) or an unlinked intergenic region (*controlΔ*) were deleted in unstable (UR-2) and stable (SR-2) caffeine-resistant isolates. **e**, A mutation in *pap1*$^+$ confers caffeine resistance in the stable isolate SR-1. Left, whole-genome sequencing of the stable isolate SR-1 revealed a 7-nucleotide insertion in *pap1*$^+$. The insertion results in a truncated Pap1 protein (Pap1-N424STOP) that lacks the nuclear export signal (NES). CRD, cysteine-rich domain. Right, Pap1-N424STOP is resistant to caffeine. The 7-nucleotide insertion identified in SR-1 was introduced into the *pap1*$^+$ gene of wild-type cells (Pap1-N424STOP) and caffeine resistance assessed. *hba1Δ* and SR-1 cells were used as positive controls. Experiments in **b**–**d** and **e**, right, were independently repeated at least twice with similar results.

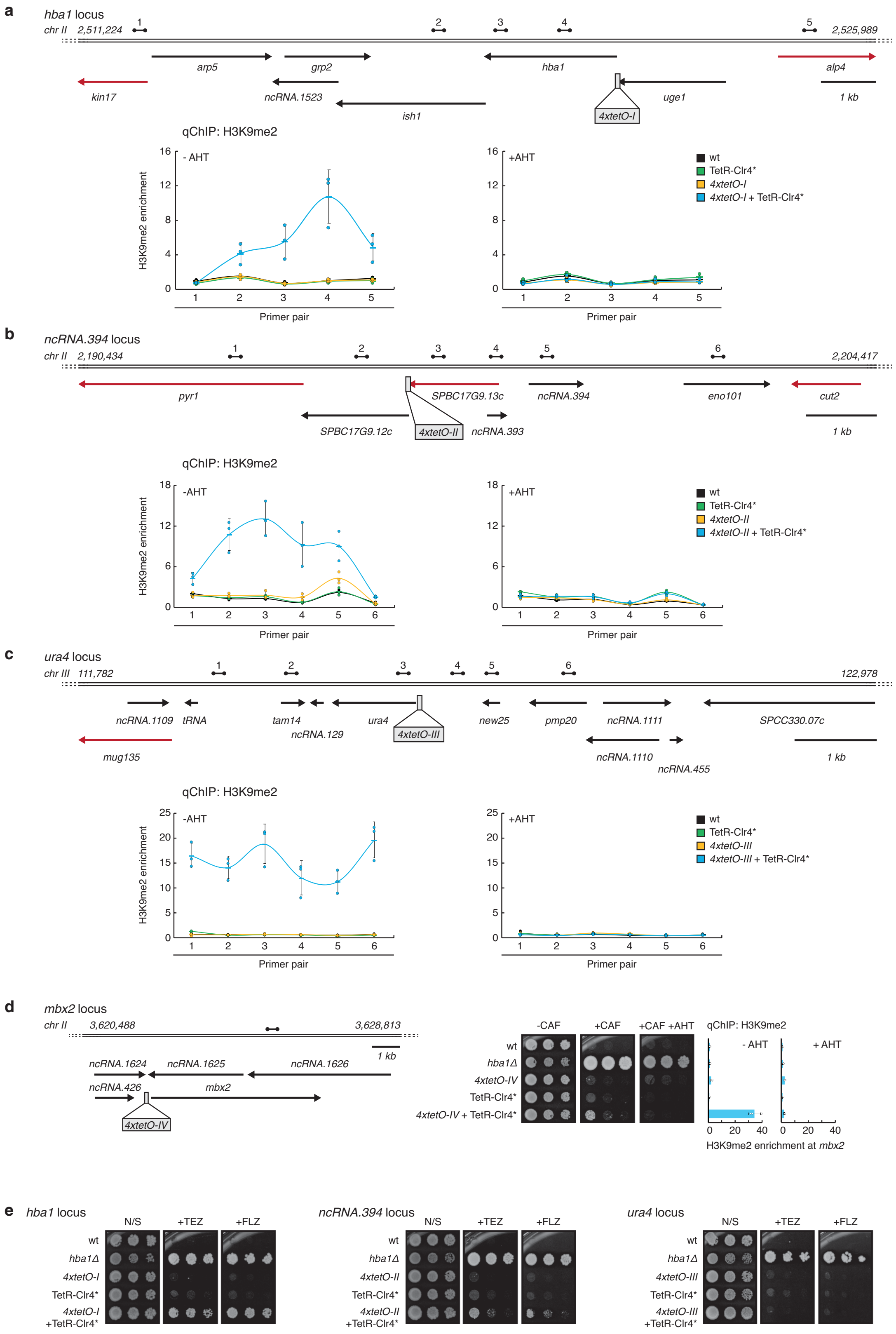


Extended Data Fig. 2 | See next page for caption.

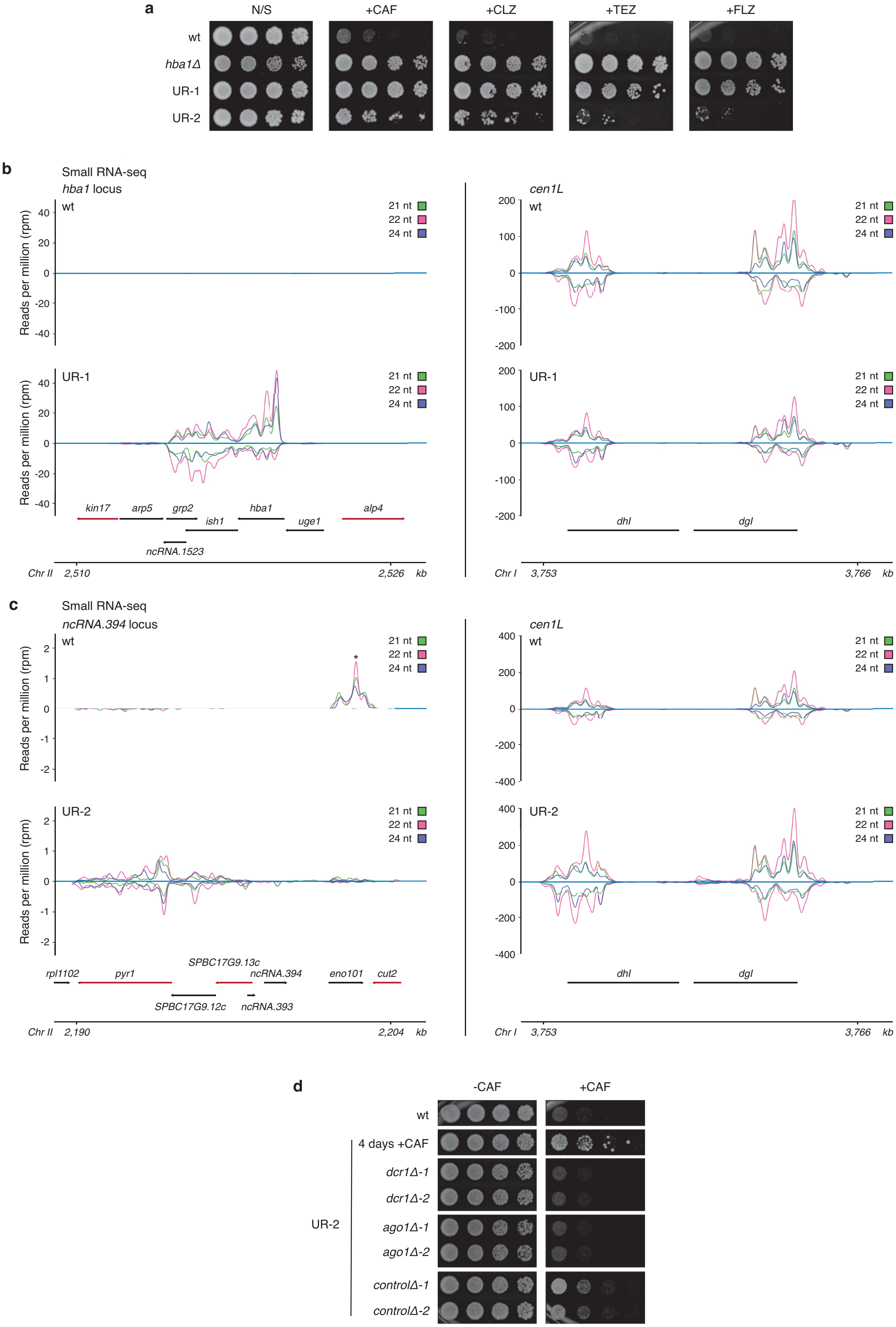
Extended Data Fig. 2 | Unstable (UR) caffeine-resistant isolates are bona fide epimutants. a–**e**, Genetic changes (*clr5-Q264STOP meu27-S100Y*) found in 4 of 30 unstable isolates do not contribute to the caffeine-resistant phenotype or cause the formation of ectopic heterochromatin. **a**, Whole-genome sequencing of unstable isolates UR-1, UR-3, UR-5 and UR-7 revealed single-nucleotide polymorphisms (SNPs) in *clr5*$^{+}$ (*clr5-Q264STOP*) and in *meu27*$^{+}$ (*meu27-S100Y*). **b**, Left, schematic of experiment to determine whether *clr5-Q264STOP meu27-S100Y* cells form more caffeine-resistant colonies than wild-type cells. Wild-type (wt) and *clr5-Q264STOP meu27-S100Y* cells were plated on +CAF medium (10^5 cells per plate, 20 plates per strain). Caffeine-resistant colonies were counted after 7 d. Right, *clr5-Q264STOP meu27-S100Y* form a similar number of caffeine-resistant colonies to wt cells. Data are mean of 20 technical replicates. *P* value from a two-tailed Student's *t*-test is indicated. **c**, *clr5-Q264STOP meu27-S100Y* cells are not resistant to caffeine. *clr5-Q264STOP meu27-S100Y* cells were serially diluted and spotted on −CAF and +CAF plates to assess caffeine resistance. *hba1Δ* cells served as a positive control. Experiment was independently repeated at least twice with similar results. **d**, Genome-wide H3K9me2 ChIP-seq enrichment in wt and *clr5-Q264STOP meu27-S100Y* cells. Data are represented as relative fold enrichment over input. **e**, H3K9me2 ChIP-seq enrichment at known heterochromatin islands detected in *epe1*Δ cells[9] in wt and *clr5-Q264STOP meu27-S100Y* cells. Data are represented as relative fold enrichment over input. **f**, Gene transcript levels within and flanking ectopic heterochromatin islands in individual isolates. See Fig. 2b. Data are mean ± s.d. from three biological replicates. *P* values <0.05 from a two-tailed Student's *t*-test are indicated.



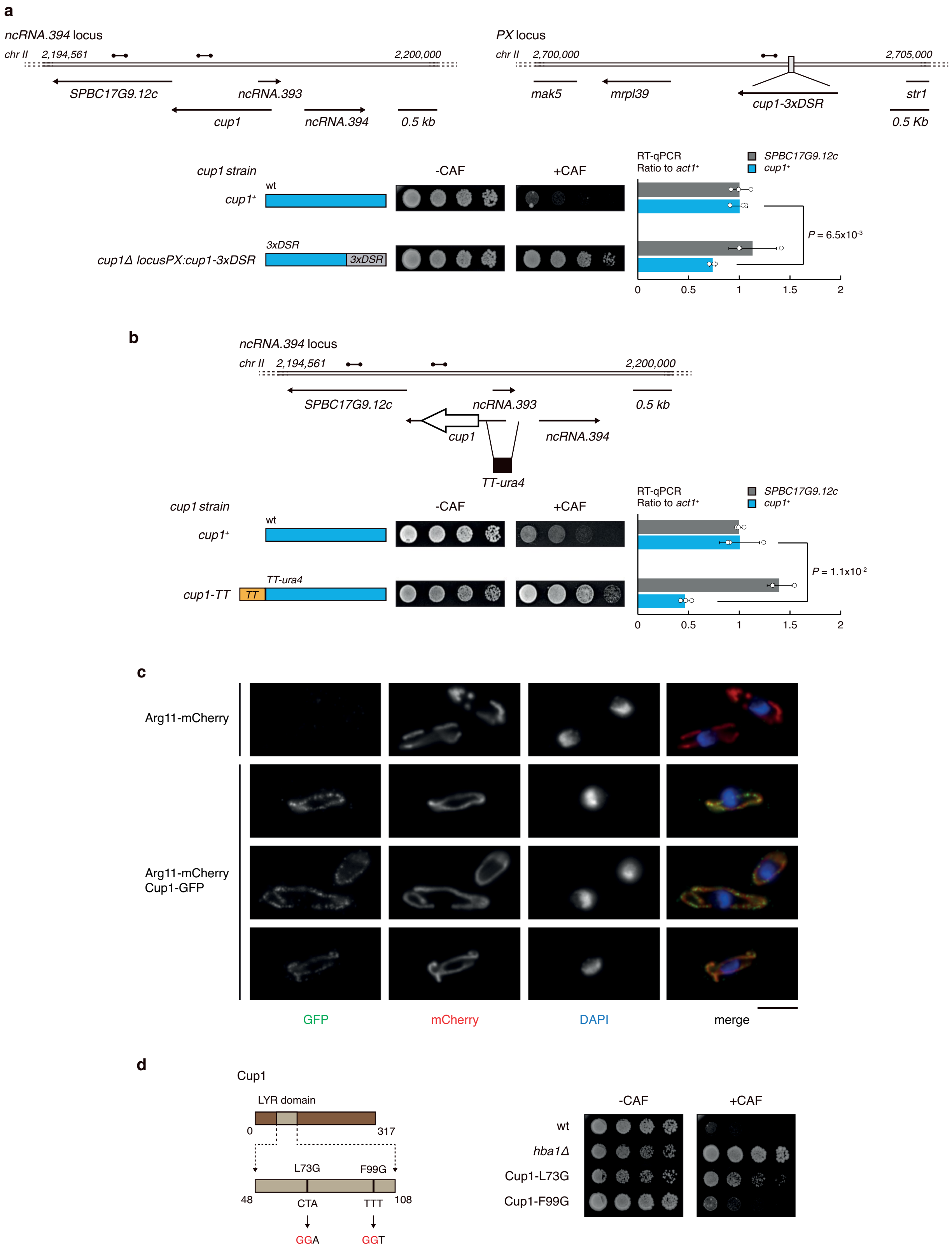
Extended Data Fig. 3 | Twenty-four of 30 unstable (UR) caffeine-resistant isolates display an ectopic heterochromatin island over the *ncRNA.394* locus. a, H3K9me2 ChIP-seq enrichment at the *ncRNA.394* locus in individual isolates (left, coverage tracks; right, heatmaps). Data are represented as relative fold enrichment over input. Relevant genes within and flanking ectopic heterochromatin islands are indicated. Red arrows indicate essential genes. Dumbbells indicate primer pairs used in **b**, **c** and **e**. **b**, Quantitative chromatin immunoprecipitation (qChIP) of H3K9me2 levels on *ncRNA.394* in individual isolates. Primer pairs used are indicated in **a** (*ncRNA.394*, primer pair 5). **c**, *SPBC17G9.13c*$^+$ gene transcript levels in individual isolates. *P* values from a two-tailed Student's *t*-test are indicated. Primer pairs used are indicated in **a** (*SPBC17G9.13c*$^+$, primer pair 3). **d**, Deletion of *ncRNA.394* or non-essential adjacent genes does not result in caffeine resistance. Experiment was independently repeated at least twice with similar results. **e**, qChIP of H3K9me2 levels at the *ncRNA.394* locus in UR-2 cells. UR-2 cells were grown in the absence (−CAF) or presence (+CAF) of caffeine overnight or in the absence of caffeine for 14 d (+14day/−CAF). Primer pairs used are indicated in **a**. In **b**, **c**, **e**, data are mean ± s.d. from three biological replicates.



Extended Data Fig. 4 | Forced synthetic heterochromatin targeting to the identified loci is sufficient to drive caffeine resistance in wild-type cells. **a**–**c**, Quantitative chromatin immunoprecipitation (qChIP) of H3K9me2 levels in wild-type (wt) cells harbouring *4xtetO* binding sites at the identified ectopic heterochromatin loci (or *ura4* as control) and expressing TetR–Clr4* in the absence or presence of AHT. **a**, *hba1* locus. **b**, *ncRNA.394* locus. **c**, *ura4* locus. Data are mean ± s.d. from three biological replicates. Dumbbells indicate primer pairs used. Red arrows indicate essential genes. **d**, Forced synthetic heterochromatin targeting to the *mbx2* locus is sufficient to drive caffeine resistance in wt cells. qChIP of H3K9me2 levels in wt cells harbouring *4xtetO* binding sites at the *mbx2* ectopic heterochromatin locus and expressing TetR–Clr4* in the absence or presence of AHT. Data are mean ± s.d. from three biological replicates. Dumbbells indicate primer pairs used. **e**, Strains from **a**–**c** were assessed for resistance to the antifungal agents tebuconazole (+TEZ) and fluconazole (+FLZ). Experiments were independently repeated at least twice with similar results.

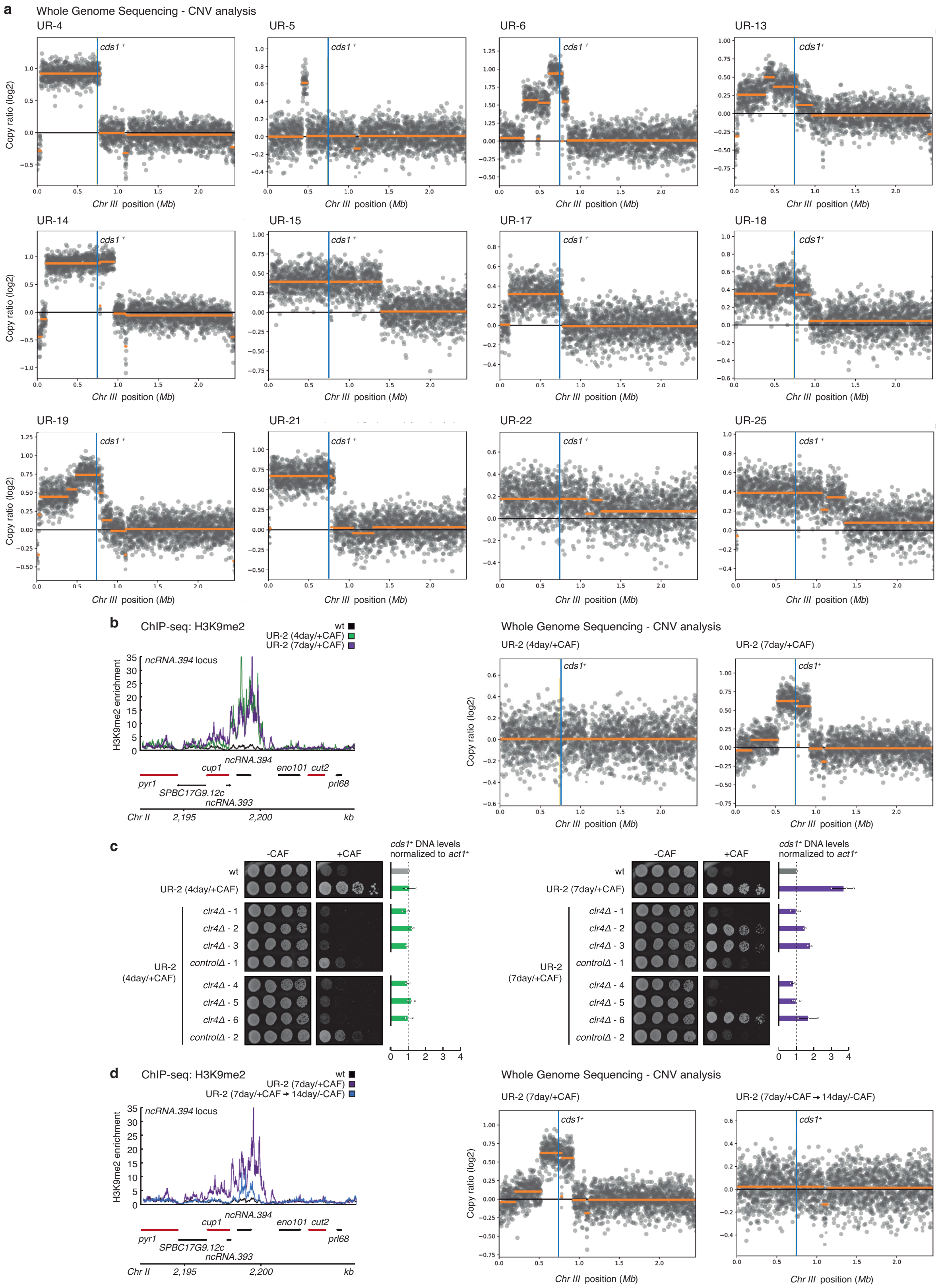


Extended Data Fig. 5 | Unstable (UR) caffeine-resistant isolates show cross-resistance to antifungals and siRNA generation at ectopic heterochromatin islands. **a**, Unstable caffeine-resistant isolates UR-1 and UR-2 were serially diluted and spotted on non-selective (N/S), caffeine (+CAF), clotrimazole (+CLZ), tebuconazole (+TEZ) and fluconazole (+FLZ) medium to assess resistance. Experiment was independently repeated at least twice with similar results. **b**, **c**, Left, small RNA sequencing detects siRNAs (21–24 nucleotides) homologous to ectopic heterochromatin islands in UR-1 (**b**, *hba1* locus) and UR-2 (**c**, *ncRNA.394* locus) compared to wild-type (wt) cells. Right, siRNAs mapping to pericentromeric *dgl/dhl* repeats of chromosome I shown as control. Sequencing was performed once. *, transcripts mapping to the highly expressed gene *eno101*$^+$ in euchromatic wild-type conditions (note these are unidirectional RNAs and not siRNAs). **d**, Caffeine resistance depends on RNAi. *dcr1*$^+$ (*dcr1Δ*), *ago1*$^+$ (*ago1Δ*) or an unlinked intergenic region (*controlΔ*) were deleted in UR-2 cells. Experiments were independently repeated at least twice with similar results.



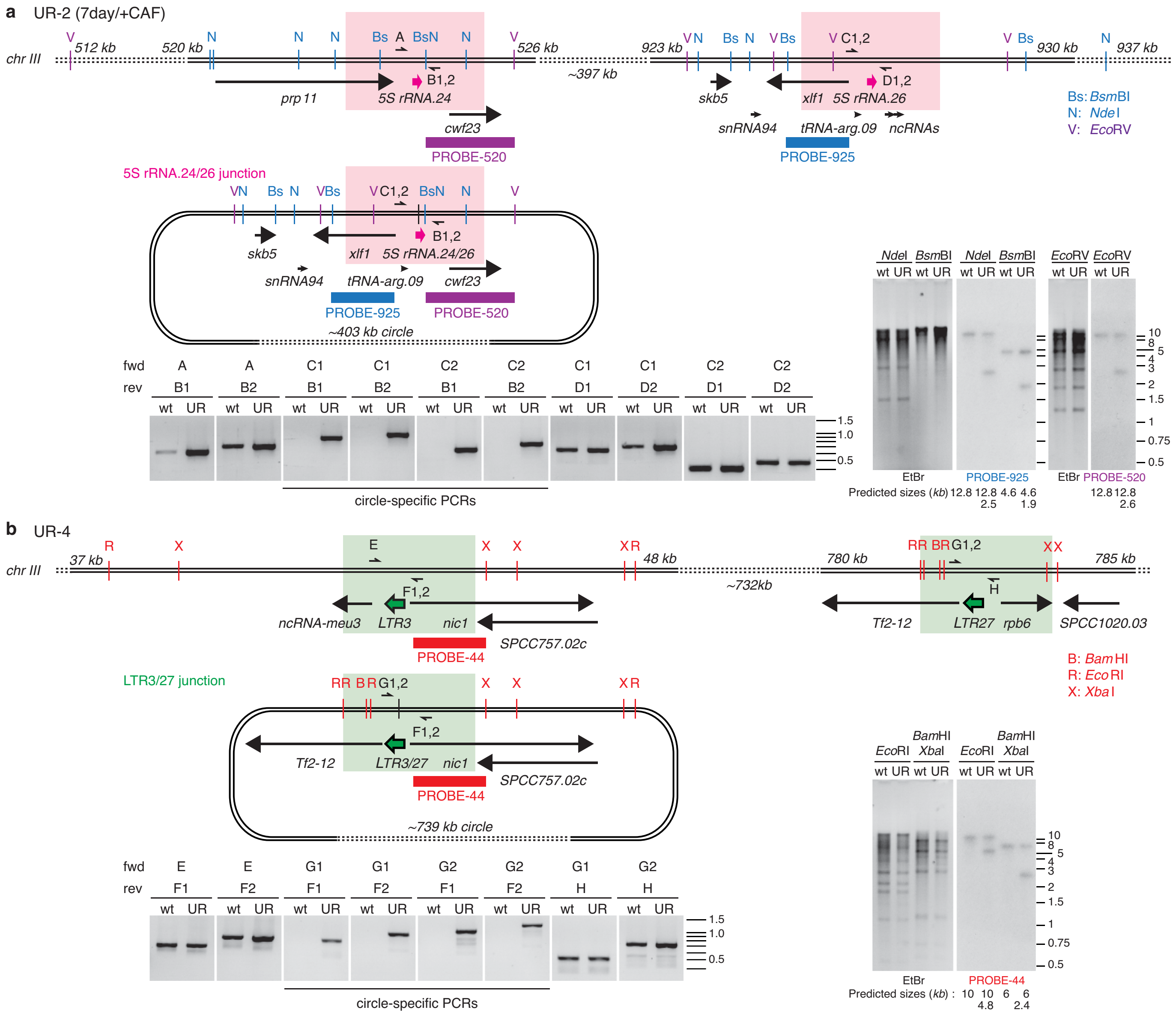
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Decreased *cup1*$^{+}$ transcript levels or Cup1 LYR-domain mutation results in caffeine resistance. a, An additional copy of *cup1*$^{+}$ with 3× determinant of selective removal (DSR) motifs fused to its 3′ untranslated region was inserted at an intergenic region (*LocusPX:cup1-3xDSR*). Bottom left, after deletion of endogenous *cup1*$^{+}$, cells expressing only *cup1-3xDSR* were assessed for caffeine resistance. Bottom right, transcript levels of *cup1*$^{+}$ and *SPBC17G9.12c*$^{+}$ (as control) in *cup1Δ locusPX:cup1-3xDSR* cells compared to wild-type. Data are mean ± s.d. from three biological replicates. *P* value from a two-tailed Student's *t*-test is indicated. Dumbbells indicate primer pairs used. **b**, The 144-bp transcriptional terminator site from *ura4*$^{+}$ was inserted in place of part of the putative *cup1*$^{+}$ promoter (*cup1-TT*). Bottom left, cells were assessed for caffeine resistance. Bottom right, transcript levels of *cup1*$^{+}$ and *SPBC17G9.12c*$^{+}$ (as control) in *cup1-TT* cells compared to wild-type. Data are mean ± s.d. from three biological replicates. *P* value from a two-tailed Student's *t*-test is indicated. Dumbbells indicate primer pairs used. **c**, Cup1 localizes to mitochondria. Cells expressing either untagged Cup1 (top row) or Cup1–GFP (bottom three rows) were fixed and processed for immunofluorescence with anti-GFP antibody and Alexa-488 secondary antibody and DNA was stained with DAPI. The mitochondrial protein Arg11–mCherry served as a positive control for mitochondrial localization. All images in the green channel (Cup1–GFP) are scaled relative to each other, as are those in the red channel (Arg11–mCherry); DAPI images are autoscaled. Bar, 5 µm. **d**, Point mutations (L73G and F99G) were introduced in the LYR domain of Cup1 and cells were assessed for caffeine resistance. Mutations were designed based on *Phyre2* tool analysis. *hba1Δ* cells were used as positive control. Experiments in **c** and **d** were independently repeated at least twice with similar results.

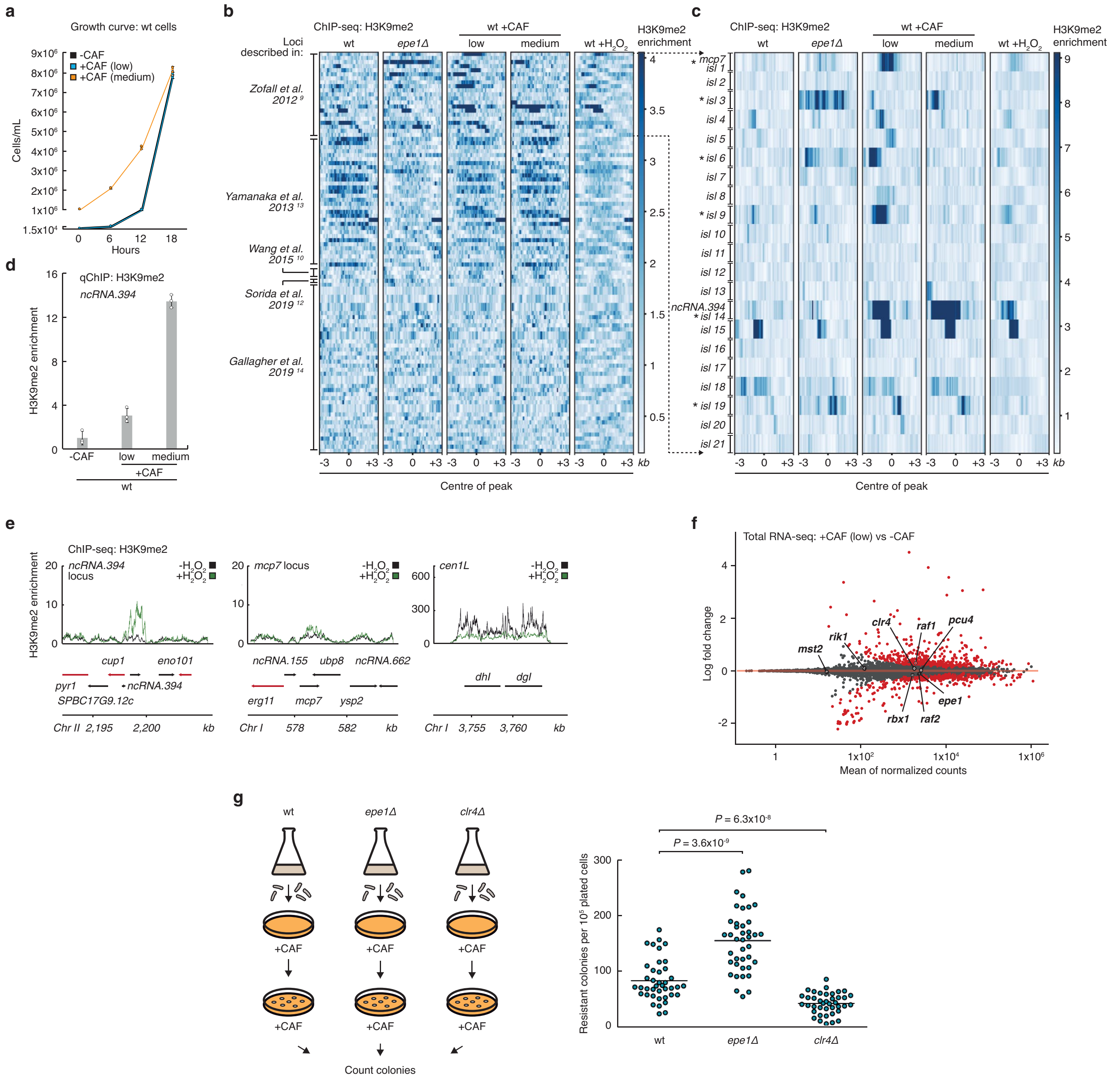


Extended Data Fig. 7 | See next page for caption.

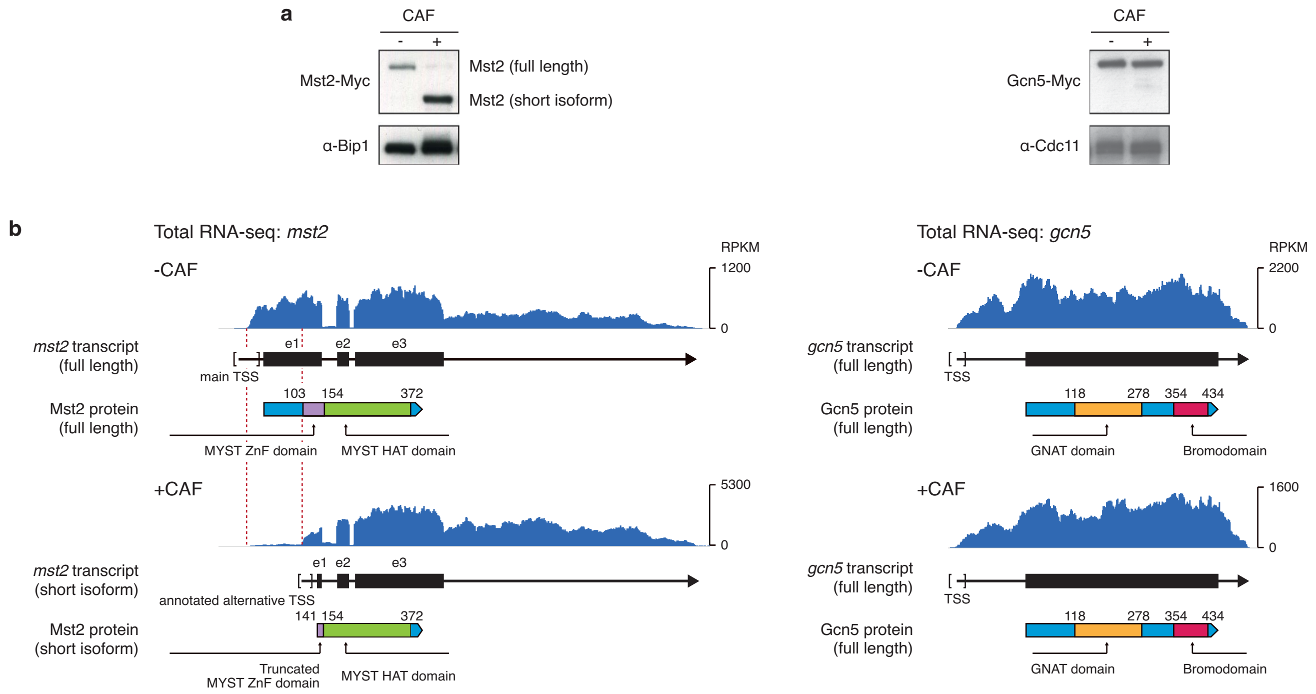
Extended Data Fig. 7 | CNV analysis reveals a partial duplication of chromosome III in 12 of 30 unstable (UR) caffeine-resistant isolates. **a**, Chromosome III coverage plots with overlaid segments in UR isolates showing partial duplication of chromosome III. Location of *cds1*$^+$ is highlighted. Wild-type ChIP-seq input data were used as the reference. **b**–**d**, Epigenetic changes preceded genetic changes (CNV) in unstable caffeine-resistant isolate UR-2. **b**, H3K9me2 ChIP-seq enrichment at the *ncRNA.394/cup1* locus (left) and chromosome III coverage plots with overlaid segments (right) in UR-2 (4day/+CAF) cells and following their prolonged growth on +CAF for an additional 3 d (7day/+CAF). Wild-type ChIP-seq input data were used as the reference for CNV analysis. **c**, *clr4*$^+$ (*clr4Δ*) or an unlinked intergenic region (*controlΔ*) were deleted in UR-2 cells (4day/+CAF) and UR-2 (7day/+CAF). All (6/6) UR-2 (4day/+CAF) *clr4Δ* transformants lost resistance to caffeine, whereas only 50% (3/6, transformants 1, 4 and 5) UR-2 (7day/+CAF) lost resistance to caffeine. Experiments were independently repeated at least twice with similar results. *cds1*$^+$ DNA levels in extracted genomic DNA were assessed by qPCR. Data are mean ± s.d. from three biological replicates. **d**, H3K9me2 ChIP-seq enrichment at the *ncRNA.394/cup1* locus (left) and chromosome III coverage plots with overlaid segments (right) in UR-2 (7day/+CAF) cells and following their prolonged growth on non-selective medium for 14 days (7day/+CAF→14day/−CAF). Wild-type ChIP-seq input data were used as the reference for CNV analysis.



Extended Data Fig. 8 | CNV of chromosome III corresponds to extrachromosomal circular DNA (eccDNA). a, **b**, Junctions of putative extrachromosomal circles were identified at repetitive sequences by inspection of CNV plots for UR-2 (7day/+CAF) (**a**) and UR-4 (**b**). In maps and lower panels, positions of *5S rRNA.24* and *5S rRNA.26* (pink arrows), *LTR3* and *LTR27* (green arrows) and flanking genes are indicated. PCR primers (half arrows) flanking *5S rRNA.24* (A (forward); B1,2 (reverse)) and *5S rRNA.26* (C1,2; D1,2) were used to amplify products from wild-type (wt) and UR-2 (7day/+CAF) ChIP input samples, along with primer combinations (C1,2; B1,2) specific for the putative circle junctions (vertical black lines). Primers flanking *LTR3* (E; F1,2) and *LTR27* (G1,2; H) were used to amplify products from wild-type and UR-4 ChIP input samples, along with primer combinations (G1,2; F1,2) specific for the putative circle junction. Shaded boxes indicate primer locations and predicted circle junctions (pink: *5S rRNA.24/26*, green: *LTR3/27*). Right, restriction enzyme-digested genomic DNA isolated from wild-type (wt), UR-2 (7day/+CAF) and UR-4 was separated on an ethidium bromide (EtBr)-containing gel followed by Southern analysis using the indicated probes (925, blue; 520, purple; 44, red). Relevant restriction enzyme sites are indicated. Predicted sizes of hybridizing fragments and DNA size markers are indicated (kb). PCR experiments were independently repeated at least twice with similar results. For gel source data, see Supplementary Fig. 1b.



Extended Data Fig. 9 | The heterochromatin profile of low-caffeine-treated wild-type cells resembles that of untreated *epe1Δ* cells. **a**, Growth of cells in caffeine. Wild-type (wt) cells were grown in the presence of low (7 mM) or medium (14 mM) caffeine for 18 h. Cell number was counted every 6 h. Note: a larger inoculum was used for 14 mM caffeine culture to obtain an equivalent final number of cells. Data are mean ± s.d. from three biological replicates. Cells from the 18-h time point were used for **d**. **b**, **c**, H3K9me2 ChIP-seq enrichment at previously-detected facultative heterochromatin loci (described in ref. [9] (**b** and **c**), ref. [13] (**b**), ref. [10] (**b**), ref. [12] (**b**) and ref. [14] (**b**)), in wt cells treated with low or medium dose of caffeine or low dose (1 mM) of H_2O_2, compared to untreated *epe1Δ* and wt cells. Data are represented as relative fold enrichment over input. A subset of facultative heterochromatin loci detected in untreated *epe1Δ* cells (refs. [9,10,12]) was detected in low-caffeine-treated wt cells. Asterisks in **c** indicate loci with similar H3K9me2 patterns in low-caffeine-treated wt cells and untreated *epe1Δ* cells, but not untreated wt cells. Facultative heterochromatin loci formed in the absence of the exosome (ref. [13]) or in wt cells grown at 18 °C (ref. [14]) were not detected in wt cells treated with low or medium caffeine or low H_2O_2. **d**, Quantitative ChIP (qChIP) of H3K9me2 levels on *ncRNA.394/cup1* in wt cells following 18 h exposure to low or medium caffeine. H3K9me2 levels were normalized to *S. octosporus* spike-in control. Data are mean ± s.d. from three biological replicates. **e**, H3K9me2 ChIP-seq enrichment at *ncRNA.394/cup1* and *mcp7* loci (or at pericentromeric *dgl/dhl* repeats of chromosome I as control) in wt cells following 18 h exposure to low H_2O_2. Data are represented as relative fold enrichment over input. Red arrows indicate essential genes. Lower levels of H3K9me2 at pericentromeric repeats upon H_2O_2 treatment may be due to H_2O_2-specific regulation of limiting heterochromatin factors at this locus. **f**, *epe1*$^+$ RNA levels do not change upon caffeine treatment. Total RNA-seq of wt cells treated with low caffeine. Transcripts encoding components of the Clr4 H3K9 methyltransferase CLRC complex (*clr4*$^+$, *rik1*$^+$, *raf1*$^+$, *raf2*$^+$, *pcu4*$^+$ and *rbx1*$^+$) and the antisilencing factors *epe1*$^+$ and *mst2*$^+$ are highlighted. Experiment was independently repeated twice with similar results. **g**, *epe1Δ* cells display increased resistance to caffeine. Left, schematic of experiment. Wild-type, *epe1Δ* and *clr4Δ* cells were plated on +CAF medium (10^5 cells/plate, 40 plates/strain). Caffeine-resistant colonies were counted after 7 d. Right, compared to wt cells, *epe1Δ* forms more, whereas *clr4Δ* forms fewer, caffeine-resistant colonies. Note that the total number of resistant colonies also includes genetic mutants. Data are mean from forty technical replicates. *P* values from a two-tailed Student's *t*-test are indicated.



Extended Data Fig. 10 | A shortened version of the anti-silencing factor Mst2 is produced upon exposure to caffeine. a, Western analysis of Mst2-13xMyc (left) and Gcn5-13xMyc (as HAT control, right) before and after caffeine treatment (medium concentration, 14 mM). Tagged proteins are expressed from their endogenous loci. Loading controls: left, Bip1; right, Cdc11. Experiments were independently repeated at least twice with similar results. For gel source data, see Supplementary Fig. 1c. **b**, Total RNA-seq for *mst2* (left) and *gcn5* (as HAT control, right) of untreated wild-type cells (top) or wild-type cells treated with medium caffeine concentration (bottom). Diagrams illustrate *mst2* and *gcn5* transcripts and predicted protein domains. Reads are normalized to RPKM. Red dashed lines indicate the region of full length *mst2* transcript absent from the short isoform. The MYST zinc finger (ZnF) domain, required for *S. cerevisiae* Esa1 acetyltransferase activity[29], is truncated in the short isoform of Mst2. The alternative *mst2* TSS used in caffeine conditions was previously annotated[28]. Experiment was independently repeated twice with similar results.

nature research

Corresponding author(s): Robin Allshire

Last updated by author(s): Jun 4, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

- Epson Connect software (Epson) for serial dilution assay picture acquisition
- Metamorph software (v7) (Universal Imaging Corporation) for cytology acquisition

Data analysis

- Trimmomatic (v0.35)
- Bowtie2 (v2.3.3)
- Samtools (v1.3.1)
- picard-tools (v2.1.0)
- deepTools (v2.0)
  - BamCompare (SES mode)
  - computeMatrix
  - plotHeatmap
- MACS2 (v2.1.1)
- IGV (v2.3.90)
- GATK HaplotypeCaller
- CNVkit (-wgs mode)
- Variant Effect Predictor (Ensembl)
- STAR (v2.2.1)
- Bioconductor (R):
  - Sushi (v1.22)
  - Rsamtools (v2.0.3)
  - GenomicFeatures (v1.36.4)
  - DESeq2 (v.1.24)
- Cutadapt (v1.17)

- SCRAM

- The complete Workflow Description Language (WDL) pipeline script used for ChIP-seq and variation analyses is available at: https://github.com/SitoTorres/Torres-Garcia-et-al.-2019

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw and processed reads from sequencing experiments are available at GEO with accession number GSE138436.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical measures were used to determine sample size. Sample sizes indicated below are considered standard in the field and were selected to ensure robust and statistically significant comparisons. For ChIP-qPCR and RT-qPCR experiments, biological triplicates (samples independently cultured for the experiment) were used. 2 biological replicates (samples independently cultured for the experiment) were performed for RNA-seq experiments. 1 small RNA-seq experiment was performed for each strain. 1 ChIP-seq experiment was performed for each strain and results were confirmed by ChIP-qPCR. For experiments performed to test whether specific genetic mutant strains form more caffeine-resistant colonies than wild-type cells, technical replicates were used to derive statistics. Equal volumes from a culture from each strain were plated on 20 or 40 (indicated in figure legend) caffeine-containing media plates. After 7 days, the number of caffeine-resistant colonies on each plate was counted. Sample sizes are provided in the figure legend.
Data exclusions	There were no data exclusions
Replication	Findings were reliably reproduced. For ChIP-qPCR and RT-qPCR, data are mean +/- standard deviation from 3 biological replicates. ChIP-seq experiments were performed once but results were confirmed by ChIP-qPCR. Serial dilution growth assays were repeated at least twice on different days with similar results.
Randomization	No randomization was required because the results of physical measurements of biomolecules, phenotypic analysis (e.g., drug resistance test) or sequencing of nucleic acid libraries are not affected by sample randomization.
Blinding	No blinding was required because the results of physical measurements of biomolecules, phenotypic analysis (e.g., drug resistance test), or sequencing of nucleic acid libraries are not affected by the researchers knowledge of sample identities.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

- Anti-H3K9me2 - Mouse monoclonal 5.1.1 - for H3K9me2 ChIP-seq and ChIP-qPCR. Kindly provided by Takeshi Urano. 1 ug per ChIP-qPCR. 3 ug per ChIP-seq.
- Anti-GFP - Invitrogen - A11122 - for GFP ChIP-qPCR and cytology. 2 ug per ChIP-qPCR. 1:500 for cytology. Lot# 2083201.
- Anti-rabbit - Invitrogen - A21441 - as secondary antibody for cytology. 1:1000. Lot# 1003212.
- Anti-FLAG-HRP - Sigma - A8591 - for western analysis. 1:5000. Lot# SLCF0816
- Anti-Myc - Cell Signalling - 9B11 - for western analysis. 1:1000. Lot# 24
- Anti-alpha-tubulin - for western analysis. Kindly provided by Keith Gull. 1:15000.
- Anti-mouse - Sigma - A4416 - as secondary antibody for western analysis. 1:10000. Lot# SLCD0197
- Anti-Bip1 - for western analysis. Lab stock. 1:1000.
- Anti-rabbit - Sigma - A6154 - as secondary antibody for western analysis. 1:10000. Lot# SLCD6835
- Anti-Cdc11 - for western analysis. Kindly provided by Ken Sawin. 1:1000.
- Anti-sheep - Abcam - ab6900 - as secondary antibody for western analysis. 1:10000.

Validation

- Mouse mAb 5.1.1: Raised in Urano Lab, validated in Nakagawachi et al (2003) Oncogone 22, 8835. Additionally, this antibody has been validated in our lab using a strain lacking the H3K9 methyltransferase.
- Anti-a-tubulin raised in Gull lab, validated in Woods et al (1989) J. Cell. Sci. 93 (Pt 3), 491–500.
- Anti-Bip1 raised and validated in Pidoux & Armstrong (1993) J. Cell. Sci. 105 (Pt 4), 1115–1120.
- Anti-Cdc11 utilized in Tong et al (2019) Nat. Commun. 10, 2343.
- All other antibodies have been extensively used for ChIP, western and cytology analyses in our laboratory and have been validated using no tag (GFP/FLAG/Myc) controls. For previous studies where these antibodies have been used see Tong et al. (2019) Nat. Commun and Bayne et al. (2010) Cell.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

All Schizosaccharomyces pombe strains used in this study are derivatives of 972 h- or other commonly used lab strains. Detailed genotypes are listed in Supplementary Table 2.

Strain number	Name	Source
143	wt	Lab stock
B4411	SR-1	This study
B4412	SR-2	This study
B4413	UR-1	This study
B4414	UR-2	This study
B4415	UR-3	This study
B4416	UR-4	This study
B4417	UR-5	This study
B4418	UR-6	This study
B4419	UR-7	This study
B4420	UR-8	This study
B4421	UR-9	This study
B4422	UR-10	This study
B4423	UR-11	This study
B4424	UR-12	This study
B4425	UR-13	This study
B4426	UR-14	This study
B4427	UR-15	This study
B4428	UR-16	This study
B4429	UR-17	This study
B4430	UR-18	This study
B4431	UR-19	This study
B4432	UR-20	This study
B4433	UR-21	This study

B4434 UR-22 This study
B4435 UR-23 This study
B4436 UR-24 This study
B4437 UR-25 This study
B4438 UR-26 This study
B4439 UR-27 This study
B4440 UR-28 This study
B4441 UR-29 This study
B4442 UR-30 This study
B4443 SR-1 clr4D - 1 This study
B4444 SR-1 clr4D - 2 This study
B4445 SR-1 NAT control - 1 This study
B4446 SR-1 NAT control - 2 This study
B4447 SR-2 clr4D - 1 This study
B4448 SR-2 clr4D - 2 This study
B4449 SR-2 NAT control - 1 This study
B4450 SR-2 NAT control - 2 This study
B4451 UR-1 clr4D - 1 This study
B4452 UR-1 clr4D - 2 This study
B4453 UR-1 NAT control-1 This study
B4454 UR-1 NAT control-2 This study
B4455 UR-2 clr4D - 1 This study
B4456 UR-2 clr4D - 2 This study
B4457 UR-2 NAT control - 1 This study
B4458 UR-2 NAT control - 2 This study
B5022 UR-2 dcr1D - 1 This study
B5023 UR-2 dcr1D - 2 This study
B5024 UR-2 ago1D - 1 This study
B5025 UR-2 ago1D - 2 This study
B4352 Pap1-N424STOP This study
B4752 Clr5-Q264STOP Meu27-S100Y This study
B4459 UR-2 +14 days -CAF This study
B4460 hba1D This study
B4461 SPBC17G9.12cD This study
B4462 ncRNA.393D This study
B4463 ncRNA.394D This study
B4464 eno101D This study
B3797 TetR-Clr4* This study
B3808 4xtetO-II This study
B3813 4xtetO-I This study
B3820 4xtetO-III This study
B4707 4xtetO-IV This study
B4465 TetR-Clr4* + 4xtetO-II This study
B4466 TetR-Clr4* + 4xtetO-I This study
B4467 TetR-Clr4* + 4xtetO-III This study
B4807 TetR-Clr4* + 4xtetO-IV This study
B4885 cup1-3xDSR This study
B5005 cup1-TT This study
B4688 Cup1-L73G This study
B4690 Cup1-F99G This study
B4567 Cup1-GFP This study
B4909 Cup1-GFP Arg11-mCh This study
B4912 Arg11-mCherry This study
B4468 UR-2 (7day/+CAF) This study
B4469 UR-2 (7day/+CAF - 14 day/-CAF) This study
B4621 epe1D This study
B2835 Epe1-GFP This study
B4958 3xFLAG-Epe1 This study
B4767 TetR-Clr4* + 4xtetO-III epe1D This study
B1008 clr4D Lab stock
B3250 S. cerevisiae Sgo1-GFP Lab stock
B3111 S. octosporus wt Lab stock
B4108 Mst2-13xMyc This study
B0505 Gcn5-13xMyc Lab stock

Authentication

All strains were generated by transformation and combined using genetic crosses. Strains were verified by the presence or absence or marker genes allowing growth on selective medium and/or by PCR to determine the presence of the desired genetic alteration and/or by DNA sequencing and/or western blotting to confirm the presence of epitope tags, as appropriate.

Mycoplasma contamination

Not applicable - Yeast strains were not tested for mycoplasma contamination, but all cultures were observed by microscopy to be free of bacterial contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified lines were used. This study used the yeast Schizosaccharomyces pombe.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

GSE138436

Files in database submission

K9me2_wt_input_P1.fastq.gz
K9me2_wt_input_P2.fastq.gz
K9me2_wt_IP_P1.fastq.gz
K9me2_wt_IP_P2.fastq.gz
K9me2_UR1_input_P1.fastq.gz
K9me2_UR1_input_P2.fastq.gz
K9me2_UR1_IP_P1.fastq.gz
K9me2_UR1_IP_P2.fastq.gz
K9me2_UR2_input_P1.fastq.gz
K9me2_UR2_input_P2.fastq.gz
K9me2_UR2_IP_P1.fastq.gz
K9me2_UR2_IP_P2.fastq.gz
K9me2_UR3_input_P1.fastq.gz
K9me2_UR3_input_P2.fastq.gz
K9me2_UR3_IP_P1.fastq.gz
K9me2_UR3_IP_P2.fastq.gz
K9me2_UR4_input_P1.fastq.gz
K9me2_UR4_input_P2.fastq.gz
K9me2_UR4_IP_P1.fastq.gz
K9me2_UR4_IP_P2.fastq.gz
K9me2_UR5_input_P1.fastq.gz
K9me2_UR5_input_P2.fastq.gz
K9me2_UR5_IP_P1.fastq.gz
K9me2_UR5_IP_P2.fastq.gz
K9me2_UR6_input_P1.fastq.gz
K9me2_UR6_input_P2.fastq.gz
K9me2_UR6_IP_P1.fastq.gz
K9me2_UR6_IP_P2.fastq.gz
K9me2_UR7_input_P1.fastq.gz
K9me2_UR7_input_P2.fastq.gz
K9me2_UR7_IP_P1.fastq.gz
K9me2_UR7_IP_P2.fastq.gz
K9me2_UR8_input_P1.fastq.gz
K9me2_UR8_input_P2.fastq.gz
K9me2_UR8_IP_P1.fastq.gz
K9me2_UR8_IP_P2.fastq.gz
K9me2_UR9_input_P1.fastq.gz
K9me2_UR9_input_P2.fastq.gz
K9me2_UR9_IP_P1.fastq.gz
K9me2_UR9_IP_P2.fastq.gz
K9me2_UR10_input_P1.fastq.gz
K9me2_UR10_input_P2.fastq.gz
K9me2_UR10_IP_P1.fastq.gz
K9me2_UR10_IP_P2.fastq.gz
K9me2_UR11_input_P1.fastq.gz
K9me2_UR11_input_P2.fastq.gz
K9me2_UR11_IP_P1.fastq.gz
K9me2_UR11_IP_P2.fastq.gz
K9me2_UR12_input_P1.fastq.gz
K9me2_UR12_input_P2.fastq.gz

K9me2_UR12_IP_P1.fastq.gz
K9me2_UR12_IP_P2.fastq.gz
K9me2_UR13_input_P1.fastq.gz
K9me2_UR13_input_P2.fastq.gz
K9me2_UR13_IP_P1.fastq.gz
K9me2_UR13_IP_P2.fastq.gz
K9me2_UR14_input_P1.fastq.gz
K9me2_UR14_input_P2.fastq.gz
K9me2_UR14_IP_P1.fastq.gz
K9me2_UR14_IP_P2.fastq.gz
K9me2_UR15_input_P1.fastq.gz
K9me2_UR15_input_P2.fastq.gz
K9me2_UR15_IP_P1.fastq.gz
K9me2_UR15_IP_P2.fastq.gz
K9me2_UR16_input_P1.fastq.gz
K9me2_UR16_input_P2.fastq.gz
K9me2_UR16_IP_P1.fastq.gz
K9me2_UR16_IP_P2.fastq.gz
K9me2_UR17_input_P1.fastq.gz
K9me2_UR17_input_P2.fastq.gz
K9me2_UR17_IP_P1.fastq.gz
K9me2_UR17_IP_P2.fastq.gz
K9me2_UR18_input_P1.fastq.gz
K9me2_UR18_input_P2.fastq.gz
K9me2_UR18_IP_P1.fastq.gz
K9me2_UR18_IP_P2.fastq.gz
K9me2_UR19_input_P1.fastq.gz
K9me2_UR19_input_P2.fastq.gz
K9me2_UR19_IP_P1.fastq.gz
K9me2_UR19_IP_P2.fastq.gz
K9me2_UR20_input_P1.fastq.gz
K9me2_UR20_input_P2.fastq.gz
K9me2_UR20_IP_P1.fastq.gz
K9me2_UR20_IP_P2.fastq.gz
K9me2_UR21_input_P1.fastq.gz
K9me2_UR21_input_P2.fastq.gz
K9me2_UR21_IP_P1.fastq.gz
K9me2_UR21_IP_P2.fastq.gz
K9me2_UR22_input_P1.fastq.gz
K9me2_UR22_input_P2.fastq.gz
K9me2_UR22_IP_P1.fastq.gz
K9me2_UR22_IP_P2.fastq.gz
K9me2_UR23_input_P1.fastq.gz
K9me2_UR23_input_P2.fastq.gz
K9me2_UR23_IP_P1.fastq.gz
K9me2_UR23_IP_P2.fastq.gz
K9me2_UR24_input_P1.fastq.gz
K9me2_UR24_input_P2.fastq.gz
K9me2_UR24_IP_P1.fastq.gz
K9me2_UR24_IP_P2.fastq.gz
K9me2_UR25_input_P1.fastq.gz
K9me2_UR25_input_P2.fastq.gz
K9me2_UR25_IP_P1.fastq.gz
K9me2_UR25_IP_P2.fastq.gz
K9me2_UR26_input_P1.fastq.gz
K9me2_UR26_input_P2.fastq.gz
K9me2_UR26_IP_P1.fastq.gz
K9me2_UR26_IP_P2.fastq.gz
K9me2_UR27_input_P1.fastq.gz
K9me2_UR27_input_P2.fastq.gz
K9me2_UR27_IP_P1.fastq.gz
K9me2_UR27_IP_P2.fastq.gz
K9me2_UR28_input_P1.fastq.gz
K9me2_UR28_input_P2.fastq.gz
K9me2_UR28_IP_P1.fastq.gz
K9me2_UR28_IP_P2.fastq.gz
K9me2_UR29_input_P1.fastq.gz
K9me2_UR29_input_P2.fastq.gz

K9me2_UR29_IP_P1.fastq.gz
K9me2_UR29_IP_P2.fastq.gz
K9me2_UR30_input_P1.fastq.gz
K9me2_UR30_input_P2.fastq.gz
K9me2_UR30_IP_P1.fastq.gz
K9me2_UR30_IP_P2.fastq.gz
K9me2_SR1_input_P1.fastq.gz
K9me2_SR1_input_P2.fastq.gz
K9me2_SR1_IP_P1.fastq.gz
K9me2_SR1_IP_P2.fastq.gz
K9me2_UR2_7_days_CAF_input_P1.fastq.gz
K9me2_UR2_7_days_CAF_input_P2.fastq.gz
K9me2_UR2_7_days_CAF_IP_P1.fastq.gz
K9me2_UR2_7_days_CAF_IP_P2.fastq.gz
K9me2_UR2_7_days_CAF_14_days_no_CAF_input_P1.fastq.gz
K9me2_UR2_7_days_CAF_14_days_no_CAF_input_P2.fastq.gz
K9me2_UR2_7_days_CAF_14_days_no_CAF_IP_P1.fastq.gz
K9me2_UR2_7_days_CAF_14_days_no_CAF_IP_P2.fastq.gz
K9me2_wt_no_treat_input_P1.fastq.gz
K9me2_wt_no_treat_input_P2.fastq.gz
K9me2_wt_no_treat_IP_P1.fastq.gz
K9me2_wt_no_treat_IP_P2.fastq.gz
K9me2_wt_7mM_CAF_input_P1.fastq.gz
K9me2_wt_7mM_CAF_input_P2.fastq.gz
K9me2_wt_7mM_CAF_IP_P1.fastq.gz
K9me2_wt_7mM_CAF_IP_P2.fastq.gz
K9me2_wt_14mM_CAF_input_P1.fastq.gz
K9me2_wt_14mM_CAF_input_P2.fastq.gz
K9me2_wt_14mM_CAF_IP_P1.fastq.gz
K9me2_wt_14mM_CAF_IP_P2.fastq.gz
K9me2_wt_1mM_H2O2_input_P1.fastq.gz
K9me2_wt_1mM_H2O2_input_P2.fastq.gz
K9me2_wt_1mM_H2O2_IP_P1.fastq.gz
K9me2_wt_1mM_H2O2_IP_P2.fastq.gz
sRNA_wt.fastq.gz
sRNA_UR1.fastq.gz
sRNA_UR2.fastq.gz
K9me2_wt_ratio.bw
K9me2_UR1_ratio.bw
K9me2_UR2_ratio.bw
K9me2_UR3_ratio.bw
K9me2_UR4_ratio.bw
K9me2_UR5_ratio.bw
K9me2_UR6_ratio.bw
K9me2_UR7_ratio.bw
K9me2_UR8_ratio.bw
K9me2_UR9_ratio.bw
K9me2_UR10_ratio.bw
K9me2_UR11_ratio.bw
K9me2_UR12_ratio.bw
K9me2_UR13_ratio.bw
K9me2_UR14_ratio.bw
K9me2_UR15_ratio.bw
K9me2_UR16_ratio.bw
K9me2_UR17_ratio.bw
K9me2_UR18_ratio.bw
K9me2_UR19_ratio.bw
K9me2_UR20_ratio.bw
K9me2_UR21_ratio.bw
K9me2_UR22_ratio.bw
K9me2_UR23_ratio.bw
K9me2_UR24_ratio.bw
K9me2_UR25_ratio.bw
K9me2_UR26_ratio.bw
K9me2_UR27_ratio.bw
K9me2_UR28_ratio.bw
K9me2_UR29_ratio.bw
K9me2_UR30_ratio.bw

K9me2_SR1_ratio.bw
K9me2_UR2_7_days_CAF_ratio.bw
K9me2_UR2_7_days_CAF_14_days_no_CAF_ratio.bw
K9me2_wt_no_treat_ratio_new.bw
K9me2_wt_7mM_CAF_ratio.bw
K9me2_wt_14mM_CAF_ratio.bw
K9me2_wt_1mM_H2O2_ratio.bw
sRNA_UR1_cen1L_locus_21nt.csv
sRNA_UR1_cen1L_locus_22nt.csv
sRNA_UR1_cen1L_locus_24nt.csv
sRNA_UR1_hba1_locus_21nt.csv
sRNA_UR1_hba1_locus_22nt.csv
sRNA_UR1_hba1_locus_24nt.csv
sRNA_UR2_cen1L_locus_21nt.csv
sRNA_UR2_cen1L_locus_22nt.csv
sRNA_UR2_cen1L_locus_24nt.csv
sRNA_UR2_ncRNA394_locus_21nt.csv
sRNA_UR2_ncRNA394_locus_22nt.csv
sRNA_UR2_ncRNA394_locus_24nt.csv
sRNA_wt_cen1L_locus_21nt.csv
sRNA_wt_cen1L_locus_22nt.csv
sRNA_wt_cen1L_locus_24nt.csv
sRNA_wt_hba1_locus_21nt.csv
sRNA_wt_hba1_locus_22nt.csv
sRNA_wt_hba1_locus_24nt.csv
sRNA_wt_ncRNA394_locus_21nt.csv
sRNA_wt_ncRNA394_locus_22nt.csv
sRNA_wt_ncRNA394_locus_24nt.csv
K9me2_clr5meu27_input_P1.fastq.gz
K9me2_clr5meu27_input_P2.fastq.gz
K9me2_clr5meu27_IP_P1.fastq.gz
K9me2_clr5meu27_IP_P2.fastq.gz
K9me2_epe1D_input_P1.fastq.gz
K9me2_epe1D_input_P2.fastq.gz
K9me2_epe1D_IP_P1.fastq.gz
K9me2_epe1D_IP_P2.fastq.gz
Total_RNAseq_wt_no_treat_1_P1.fastq.gz
Total_RNAseq_wt_no_treat_1_P2.fastq.gz
Total_RNAseq_wt_no_treat_2_P1.fastq.gz
Total_RNAseq_wt_no_treat_2_P2.fastq.gz
Total_RNAseq_wt_7mM_CAF_1_P1.fastq.gz
Total_RNAseq_wt_7mM_CAF_1_P2.fastq.gz
Total_RNAseq_wt_7mM_CAF_2_P1.fastq.gz
Total_RNAseq_wt_7mM_CAF_2_P2.fastq.gz
K9me2_clr5meu27_ratio.bw
K9me2_epe1D_ratio.bw
Total_RNAseq_wt_no_treat_1_forward.bw
Total_RNAseq_wt_no_treat_1_reverse.bw
Total_RNAseq_wt_no_treat_2_forward.bw
Total_RNAseq_wt_no_treat_2_reverse.bw
Total_RNAseq_wt_7mM_CAF_1_forward.bw
Total_RNAseq_wt_7mM_CAF_1_reverse.bw
Total_RNAseq_wt_7mM_CAF_2_forward.bw
Total_RNAseq_wt_7mM_CAF_2_reverse.bw
Total_RNAseq_wt_14mM_CAF_1_P1.fastq.gz
Total_RNAseq_wt_14mM_CAF_1_P2.fastq.gz
Total_RNAseq_wt_14mM_CAF_2_P1.fastq.gz
Total_RNAseq_wt_14mM_CAF_2_P2.fastq.gz
Total_RNAseq_wt_14mM_CAF_1_forward.bw
Total_RNAseq_wt_14mM_CAF_1_reverse.bw
Total_RNAseq_wt_14mM_CAF_2_forward.bw
Total_RNAseq_wt_14mM_CAF_2_reverse.bw

Genome browser session (e.g. UCSC)	Not applicable. Visualized data using IGV.

Methodology

Replicates

1 ChIP-seq was performed for each strain. Results were confirmed by ChIP-qPCR.

Sequencing depth

All libraries were sequenced by 75 bp paired-end reads. We did not calculate number of uniquely mapped reads, since Bowtie2 default options do not do so. Heterochromatin regions/sequences are repetitive and the number of uniquely mapped reads is not informative. Number of total reads for each fastq file are provided below:
Sample Total reads
K9me2_wt_input_P1.fastq.gz 4657568
K9me2_wt_input_P2.fastq.gz 4657568
K9me2_wt_IP_P1.fastq.gz 11558824
K9me2_wt_IP_P2.fastq.gz 11558824
K9me2_UR1_input_P1.fastq.gz 24492873
K9me2_UR1_input_P2.fastq.gz 24492873
K9me2_UR1_IP_P1.fastq.gz 23508431
K9me2_UR1_IP_P2.fastq.gz 23508431
K9me2_UR2_input_P1.fastq.gz 1470372
K9me2_UR2_input_P2.fastq.gz 1470372
K9me2_UR2_IP_P1.fastq.gz 7862615
K9me2_UR2_IP_P2.fastq.gz 7862615
K9me2_UR3_input_P1.fastq.gz 28636264
K9me2_UR3_input_P2.fastq.gz 28636264
K9me2_UR3_IP_P1.fastq.gz 22145255
K9me2_UR3_IP_P2.fastq.gz 22145255
K9me2_UR4_input_P1.fastq.gz 1742547
K9me2_UR4_input_P2.fastq.gz 1742547
K9me2_UR4_IP_P1.fastq.gz 8439033
K9me2_UR4_IP_P2.fastq.gz 8439033
K9me2_UR5_input_P1.fastq.gz 2007651
K9me2_UR5_input_P2.fastq.gz 2007651
K9me2_UR5_IP_P1.fastq.gz 8024785
K9me2_UR5_IP_P2.fastq.gz 8024785
K9me2_UR6_input_P1.fastq.gz 1842928
K9me2_UR6_input_P2.fastq.gz 1842928
K9me2_UR6_IP_P1.fastq.gz 8120241
K9me2_UR6_IP_P2.fastq.gz 8120241
K9me2_UR7_input_P1.fastq.gz 2653694
K9me2_UR7_input_P2.fastq.gz 2653694
K9me2_UR7_IP_P1.fastq.gz 8632643
K9me2_UR7_IP_P2.fastq.gz 8632643
K9me2_UR8_input_P1.fastq.gz 7639839
K9me2_UR8_input_P2.fastq.gz 7639839
K9me2_UR8_IP_P1.fastq.gz 15412350
K9me2_UR8_IP_P2.fastq.gz 15412350
K9me2_UR9_input_P1.fastq.gz 4554001
K9me2_UR9_input_P2.fastq.gz 4554001
K9me2_UR9_IP_P1.fastq.gz 11021688
K9me2_UR9_IP_P2.fastq.gz 11021688
K9me2_UR10_input_P1.fastq.gz 5176464
K9me2_UR10_input_P2.fastq.gz 5176464
K9me2_UR10_IP_P1.fastq.gz 13223029
K9me2_UR10_IP_P2.fastq.gz 13223029
K9me2_UR11_input_P1.fastq.gz 4636366
K9me2_UR11_input_P2.fastq.gz 4636366
K9me2_UR11_IP_P1.fastq.gz 12148357
K9me2_UR11_IP_P2.fastq.gz 12148357
K9me2_UR12_input_P1.fastq.gz 5608106
K9me2_UR12_input_P2.fastq.gz 5608106
K9me2_UR12_IP_P1.fastq.gz 12336535
K9me2_UR12_IP_P2.fastq.gz 12336535
K9me2_UR13_input_P1.fastq.gz 1632034
K9me2_UR13_input_P2.fastq.gz 1632034
K9me2_UR13_IP_P1.fastq.gz 7052099
K9me2_UR13_IP_P2.fastq.gz 7052099
K9me2_UR14_input_P1.fastq.gz 1638982
K9me2_UR14_input_P2.fastq.gz 1638982
K9me2_UR14_IP_P1.fastq.gz 8411534
K9me2_UR14_IP_P2.fastq.gz 8411534

K9me2_UR15_input_P1.fastq.gz 1328085
K9me2_UR15_input_P2.fastq.gz 1328085
K9me2_UR15_IP_P1.fastq.gz 7511500
K9me2_UR15_IP_P2.fastq.gz 7511500
K9me2_UR16_input_P1.fastq.gz 1574195
K9me2_UR16_input_P2.fastq.gz 1574195
K9me2_UR16_IP_P1.fastq.gz 7552346
K9me2_UR16_IP_P2.fastq.gz 7552346
K9me2_UR17_input_P1.fastq.gz 1906114
K9me2_UR17_input_P2.fastq.gz 1906114
K9me2_UR17_IP_P1.fastq.gz 8182937
K9me2_UR17_IP_P2.fastq.gz 8182937
K9me2_UR18_input_P1.fastq.gz 1684998
K9me2_UR18_input_P2.fastq.gz 1684998
K9me2_UR18_IP_P1.fastq.gz 4860918
K9me2_UR18_IP_P2.fastq.gz 4860918
K9me2_UR19_input_P1.fastq.gz 1875542
K9me2_UR19_input_P2.fastq.gz 1875542
K9me2_UR19_IP_P1.fastq.gz 7053803
K9me2_UR19_IP_P2.fastq.gz 7053803
K9me2_UR20_input_P1.fastq.gz 1755882
K9me2_UR20_input_P2.fastq.gz 1755882
K9me2_UR20_IP_P1.fastq.gz 7803452
K9me2_UR20_IP_P2.fastq.gz 7803452
K9me2_UR21_input_P1.fastq.gz 1854096
K9me2_UR21_input_P2.fastq.gz 1854096
K9me2_UR21_IP_P1.fastq.gz 7963343
K9me2_UR21_IP_P2.fastq.gz 7963343
K9me2_UR22_input_P1.fastq.gz 1534548
K9me2_UR22_input_P2.fastq.gz 1534548
K9me2_UR22_IP_P1.fastq.gz 7713816
K9me2_UR22_IP_P2.fastq.gz 7713816
K9me2_UR23_input_P1.fastq.gz 1786133
K9me2_UR23_input_P2.fastq.gz 1786133
K9me2_UR23_IP_P1.fastq.gz 7886760
K9me2_UR23_IP_P2.fastq.gz 7886760
K9me2_UR24_input_P1.fastq.gz 1623522
K9me2_UR24_input_P2.fastq.gz 1623522
K9me2_UR24_IP_P1.fastq.gz 8527474
K9me2_UR24_IP_P2.fastq.gz 8527474
K9me2_UR25_input_P1.fastq.gz 1664888
K9me2_UR25_input_P2.fastq.gz 1664888
K9me2_UR25_IP_P1.fastq.gz 8235632
K9me2_UR25_IP_P2.fastq.gz 8235632
K9me2_UR26_input_P1.fastq.gz 1674916
K9me2_UR26_input_P2.fastq.gz 1674916
K9me2_UR26_IP_P1.fastq.gz 6584663
K9me2_UR26_IP_P2.fastq.gz 6584663
K9me2_UR27_input_P1.fastq.gz 1591681
K9me2_UR27_input_P2.fastq.gz 1591681
K9me2_UR27_IP_P1.fastq.gz 7201369
K9me2_UR27_IP_P2.fastq.gz 7201369
K9me2_UR28_input_P1.fastq.gz 1700557
K9me2_UR28_input_P2.fastq.gz 1700557
K9me2_UR28_IP_P1.fastq.gz 8257483
K9me2_UR28_IP_P2.fastq.gz 8257483
K9me2_UR29_input_P1.fastq.gz 1697044
K9me2_UR29_input_P2.fastq.gz 1697044
K9me2_UR29_IP_P1.fastq.gz 8492861
K9me2_UR29_IP_P2.fastq.gz 8492861
K9me2_UR30_input_P1.fastq.gz 1539579
K9me2_UR30_input_P2.fastq.gz 1539579
K9me2_UR30_IP_P1.fastq.gz 7862112
K9me2_UR30_IP_P2.fastq.gz 7862112
K9me2_SR1_input_P1.fastq.gz 5049507
K9me2_SR1_input_P2.fastq.gz 5049507
K9me2_SR1_IP_P1.fastq.gz 11333792
K9me2_SR1_IP_P2.fastq.gz 11333792

K9me2_UR2_7_days_CAF_input_P1.fastq.gz 1590914
K9me2_UR2_7_days_CAF_input_P2.fastq.gz 1590914
K9me2_UR2_7_days_CAF_IP_P1.fastq.gz 8276251
K9me2_UR2_7_days_CAF_IP_P2.fastq.gz 8276251
K9me2_UR2_7_days_CAF_14_days_no_CAF_input_P1.fastq.gz 1668325
K9me2_UR2_7_days_CAF_14_days_no_CAF_input_P2.fastq.gz 1668325
K9me2_UR2_7_days_CAF_14_days_no_CAF_IP_P1.fastq.gz 7837845
K9me2_UR2_7_days_CAF_14_days_no_CAF_IP_P2.fastq.gz 7837845
K9me2_wt_no_treat_input_P1.fastq.gz 2484101
K9me2_wt_no_treat_input_P2.fastq.gz 2484101
K9me2_wt_no_treat_IP_P1.fastq.gz 7881079
K9me2_wt_no_treat_IP_P2.fastq.gz 7881079
K9me2_wt_7mM_CAF_input_P1.fastq.gz 2606829
K9me2_wt_7mM_CAF_input_P2.fastq.gz 2606829
K9me2_wt_7mM_CAF_IP_P1.fastq.gz 9212592
K9me2_wt_7mM_CAF_IP_P2.fastq.gz 9212592
K9me2_wt_14mM_CAF_input_P1.fastq.gz 2472114
K9me2_wt_14mM_CAF_input_P2.fastq.gz 2472114
K9me2_wt_14mM_CAF_IP_P1.fastq.gz 7870676
K9me2_wt_14mM_CAF_IP_P2.fastq.gz 7870676
K9me2_wt_1mM_H2O2_input_P1.fastq.gz 6680370
K9me2_wt_1mM_H2O2_input_P2.fastq.gz 6680370
K9me2_wt_1mM_H2O2_IP_P1.fastq.gz 10697650
K9me2_wt_1mM_H2O2_IP_P2.fastq.gz 10697650
K9me2_clr5meu27_input_P1.fastq.gz 6748422
K9me2_clr5meu27_input_P2.fastq.gz 6748422
K9me2_clr5meu27_IP_P1.fastq.gz 30064927
K9me2_clr5meu27_IP_P2.fastq.gz 30064927
K9me2_epe1D_input_P1.fastq.gz 1542750
K9me2_epe1D_input_P2.fastq.gz 1542750
K9me2_epe1D_IP_P1.fastq.gz 5929968
K9me2_epe1D_IP_P2.fastq.gz 5929968

Antibodies

Mouse mAb 5.1.1 anti-H3K9me2 for H3K9me2 ChIP-seq

Peak calling parameters

```
macs2 callpeak -f BAMPE -t sample.bam -c sample.bam --broad -g 14e6 --broad-cutoff 0.05 -n sample
```

Data quality

MACS2 was used to call peaks from paired-end ChIP-seq reads

Software

- Trimmomatic (v0.35)
- Bowtie2 (v2.3.3)
- Samtools (v1.3.1)
- picard-tools (v2.1.0)
- deepTools (v2.0)
  - BamCompare (SES mode)
  - computeMatrix
  - plotHeatmap
- MACS2 (v2.1.1)
- IGV (v2.3.90)
- Bioconductor (R):
  - Sushi (v1.22)
- The complete Workflow Description Language (WDL) pipeline script used for ChIP-seq and variation analyses is available at: https://github.com/SitoTorres/Torres-Garcia-et-al.-2019

Article

Identification of the human DPR core promoter element using machine learning

https://doi.org/10.1038/s41586-020-2689-7

Received: 27 November 2019

Accepted: 16 June 2020

Long Vo ngoc[1], Cassidy Yunjing Huang[1], California Jack Cassidy[1], Claudia Medrano[1] & James T. Kadonaga[1]✉

The RNA polymerase II (Pol II) core promoter is the strategic site of convergence of the signals that lead to the initiation of DNA transcription[1–5], but the downstream core promoter in humans has been difficult to understand[1–3]. Here we analyse the human Pol II core promoter and use machine learning to generate predictive models for the downstream core promoter region (DPR) and the TATA box. We developed a method termed HARPE (high-throughput analysis of randomized promoter elements) to create hundreds of thousands of DPR (or TATA box) variants, each with known transcriptional strength. We then analysed the HARPE data by support vector regression (SVR) to provide comprehensive models for the sequence motifs, and found that the SVR-based approach is more effective than a consensus-based method for predicting transcriptional activity. These results show that the DPR is a functionally important core promoter element that is widely used in human promoters. Notably, there appears to be a duality between the DPR and the TATA box, as many promoters contain one or the other element. More broadly, these findings show that functional DNA motifs can be identified by machine learning analysis of a comprehensive set of sequence variants.

The core promoter is generally considered to be the stretch of DNA that directs the initiation of transcription of a gene; it ranges from about −40 to +40 nucleotides (nt) relative to the +1 nt transcription start site (TSS)[1–5]. The core promoter comprises DNA sequence elements such as the TATA box, initiator (Inr), motif ten element (MTE), and downstream core promoter element (DPE) (Extended Data Fig. 1a). Each of these motifs is present only at a subset of core promoters. Hence, there are no universal core promoter elements. Moreover, specific core promoter motifs can be important for enhancer–promoter specificity[6–8] and can be involved in gene networks[7,9–11].

The key DNA sequence motifs of human core promoters remain to be clarified. In focused human promoters, in which transcription initiates at a single site or a narrow cluster of sites, the TATA box is the best known core promoter element, but most human core promoters lack a TATA box[12]. In *Drosophila*, TATA-less transcription is frequently driven by the downstream MTE and DPE motifs[13–16]; however, these motifs have rarely been found in human promoters and have been thought perhaps not to exist in humans[1–3].

HARPE analysis of the downstream promoter

To decipher the downstream core promoter in humans, we generated and analysed an extensive library of promoters that contain randomized sequences in the region from +17 to +35 nt relative to the +1 nt TSS. This stretch, which we term the DPR, comprises the positions that correspond to the MTE and DPE (Fig. 1a, Extended Data Fig. 1a), which are overlapping elements in the downstream core promoter region in *Drosophila* that span multiple contact points with the transcription factor TFIID[16–19]. In previous studies, libraries of entire core promoter regions have been screened and characterized by using cell-based systems[20–24]. By contrast, here we have analysed specific segments of the core promoter in vitro and in cells, with the strategy of obtaining high coverage and carrying out machine learning analysis of the data.

In natural promoters, it can be difficult to elucidate the characteristics of a specific DNA element, such as the DPR, owing to the different promoter backgrounds in which the sequence motif is situated. To circumvent this problem, we adapted the survey of regulatory elements[23] (SuRE) and developed the HARPE method. HARPE involves the generation of around 500,000 random DPR variants in an invariant promoter cassette followed by assessment of the transcription strength (defined as the RNA tag count divided by the DNA tag count; Methods) of each variant in vitro (Fig. 1a, Extended Data Fig. 1, Supplementary Table 1). This analysis showed that most DPR sequence variants support only a low level of transcription (Fig. 1b) and that the most active DPR sequences exhibit distinct nucleotide preferences (Extended Data Fig. 1d). Moreover, hypergeometric optimization of motif enrichment (HOMER) motif discovery analysis[25] of the top 0.1% most-transcribed HARPE sequences identified a distinct motif that resembled the *Drosophila* DPE consensus sequence (RGWYGT from +28 to +32)[14] (Fig. 1c, Extended Data Fig. 1e, f). The results of HARPE are reproducible (Extended Data Fig. 1g–i) in the absence or presence of sarkosyl, which limits transcription to a single round (Extended Data Fig. 2a–d, Supplementary Discussion 1).

HARPE is a robust and versatile method

To determine the versatility of the HARPE method, we tested the assay by varying different experimental parameters. First, we compared

---

[1]Section of Molecular Biology, University of California, San Diego, La Jolla, CA, USA. ✉e-mail: jkadonaga@ucsd.edu

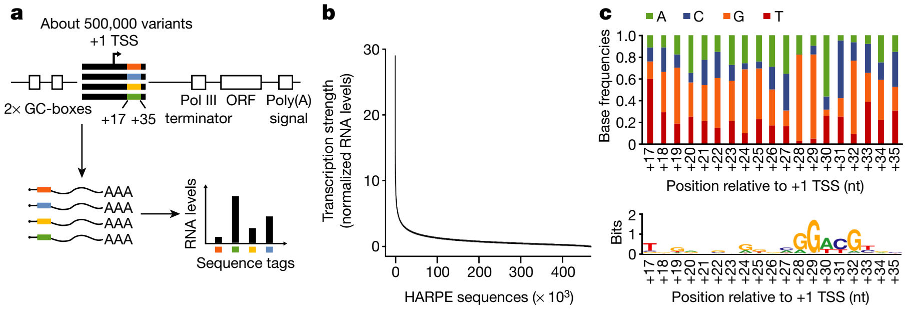


Fig. 1 | HARPE comprehensively assesses the transcriptional effect of many different DNA sequences in a specific region of the promoter. a, Schematic of HARPE for the analysis of DNA sequence variants in the DPR. The randomized segment was generated by oligonucleotide synthesis with mixed nucleotides. ORF, open reading frame. **b**, Most sequence variants exhibit low transcriptional activity. The distribution of transcription strength for each of the approximately 500,000 core promoter variants is shown. **c**, A distinct DPR sequence motif can be seen in the nucleotide frequencies of the 0.1% most transcribed DPR sequences (top) as well as in the web logo for the top HOMER motif that is identified with these sequences (bottom). All panels show a representative experiment, *n* = 2 biologically independent samples.

the results of HARPE assays that were performed with two different core promoter cassettes: SCP1m (as in Fig. 1), which is a version of the synthetic SCP1 promoter with a mutant TATA box (also known as SCP1mTATA[26]); and the human IRF1 core promoter, which lacks a TATA box and contains a DPE motif[17]. Both core promoters contain a consensus Inr sequence[12], but otherwise they share no sequence similarity. With these two different core promoter cassettes, the HARPE results were nearly indistinguishable (Fig. 2a, Extended Data Figs. 1i, 2e). In addition, we observed nearly the same results with TATA-less versus TATA-box-containing promoters (Fig. 2b, Extended Data Figs. 1i, 2e). Thus, HARPE can function consistently in different core promoter backgrounds.

Second, we investigated whether we would obtain consistent HARPE data if we randomized only a subset of the DPR rather than the entire DPR. To this end, we performed HARPE by randomization of only the MTE region (+18 to +29 nt) or only an extended DPE region (+23 to +34 nt) (Fig. 2c, Extended Data Figs. 1i, 2f). These experiments showed that randomization of subregions of the DPR yielded nucleotide preferences similar to those obtained by randomization of the entire DPR.

Third, we tested whether transcription of the HARPE promoter libraries in cells would yield results similar to those seen in vitro (Fig. 2d, Extended Data Fig. 2g). To this end, we carried out HARPE by transfection of the promoter libraries into HeLa cells and observed nucleotide preferences in the DPR that were nearly identical to those seen in vitro. Furthermore, we found a strong resemblance between HARPE data generated in vitro and in cells with the DPR sequence in the human IRF1 and TATA-box-containing SCP1 core promoter cassettes, as well as with the MTE and DPE sequences (Extended Data Fig. 2h–j). Therefore, HARPE appears to be a robust method that provides consistent data under a variety of different conditions.

HARPE analysis of the upstream TATA box

To enable the use of HARPE for the analysis of upstream promoter elements, we developed a modified version that includes linkage of each of the upstream randomized motifs with a corresponding downstream barcode (Extended Data Fig. 2k–p). We performed this analysis with randomized sequences in the region of the TATA box. We tested a long TATA region (−32 to −21 nt relative to the +1 nt TSS) and a short TATA region (−30 to −23 nt) (Extended Data Figs. 1a, 2k–p). The long-TATA analysis yielded an A/T-rich stretch that resembled that seen in natural human promoters. The short-TATA construct contained a TA dinucleotide at positions −32 and −31 that served to fix the phasing of the TATA sequence. Hence, with the short TATA construct, we observed a more distinct TATA-box-like sequence in a single register. Thus, HARPE can be used to analyse upstream as well as downstream promoter sequences.

Machine learning analysis of the HARPE data

HARPE analysis of the DPR yielded hundreds of thousands of sequence variants (Supplementary Table 1), each of which was associated with a specific transcription strength, and the data were therefore well suited for machine learning analysis. There are many different methods for supervised learning, and we found SVR[27,28] to be an effective and straightforward approach for the analysis of the HARPE data.

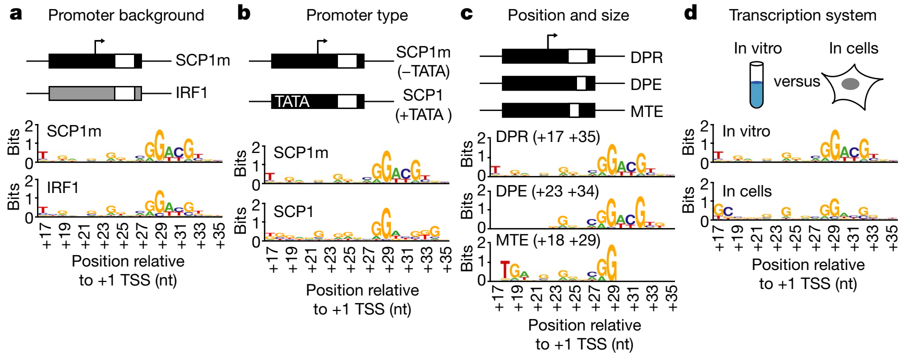


Fig. 2 | HARPE yields consistent data under different conditions. The top HOMER motifs obtained from the 0.1% most active sequences are shown. **a**, HARPE of the DPR with two different promoter cassettes: SCP1 lacking a TATA box (SCP1m) and the human IRF1 core promoter (in vitro transcription). **b**, HARPE of the DPR with a TATA-less promoter (SCP1m) and a TATA-box-containing promoter (SCP1) in vitro. **c**, HARPE of the DPR (+17 to +35 nt), DPE (+23 to +34 nt), and MTE (+18 to +29 nt) motifs with the SCP1m promoter in vitro. **d**, HARPE of the DPR in the SCP1m promoter transcribed in vitro or in cells. All panels show a representative experiment, *n* = 2 biologically independent samples.

In the SVR analysis of the DPR, we started with 468,069 sequence variants, each of which had a known transcriptional strength (Fig. 3a). We set aside 7,500 sequences that represented the full range of observed transcription strengths (test sequences) for later testing of the SVR. Next, we trained the SVR with 200,000 sequences (Extended Data Fig. 3a) and performed grid search and cross validation to identify optimal hyperparameter values and to establish the stability of the model (Extended Data Fig. 3b–d). The resulting SVR model that was generated from the biochemical (in vitro transcription) data was termed SVRb.

The SVRb model was then able to provide a numerical value for the predicted transcription strength of any DNA sequence. First, we found an excellent correlation ($\rho = 0.90$) between the predicted SVRb scores and the observed transcription strengths of independent test sequences (Fig. 3b, Extended Data Fig. 3e). Second, we generated and analysed a separate high-quality, low-complexity HARPE dataset of DPR variants (Extended Data Fig. 3f–i), and saw an excellent correlation ($\rho = 0.96$) between the predicted SVRb scores and the observed transcription strengths (Fig. 3c). Third, we individually transcribed 16 promoters with a range of SVRb scores (Extended Data Fig. 4). These experiments revealed an excellent correlation ($\rho = 0.89$–0.95) between the predicted SVRb scores and the transcriptional activities of the individual sequences tested in vitro and in cells (Fig. 3d, Extended Data Fig. 4). It is also important to note that sequence variants with an SVRb score of two or more typically have at least sixfold-higher activity than inactive sequences (comparison of median values in the two groups; Extended Data Fig. 5a–c). Thus, an SVRb score of two or more is likely to reflect an active DPR. Last, performance assessment of SVRb revealed that it reliably predicts active DPR sequences (Extended Data Fig. 5d–r).

The data thus indicate that SVRb provides an accurate model for the DPR. Furthermore, we observed that SVRb, which was created with the SCP1m promoter cassette, correlated well with an SVRIRF1 model that was generated with HARPE data for the DPR with the human IRF1 promoter cassette ($\rho = 0.87$) (Extended Data Fig. 6a, b). We also saw a good correlation between SVRb (for the DPR in a TATA-less background) and SVRSCP1, which was generated with HARPE data for the DPR with the SCP1 (TATA-containing) promoter cassette ($\rho = 0.80$) (Extended Data Fig. 6c–e). Hence, the combination of HARPE and SVR analysis yields similar SVR models with different promoter backgrounds.

SVR models versus consensus sequences

To test the utility of an SVR model relative to a consensus sequence, we compared DPR sequences that were obtained by a standard consensus approach to the scores predicted by SVRb. First, we identified the DPE-like RGWYGT consensus sequence (from +28 to +33 nt) in the top 0.1% most active HARPE variants (Fig. 1c, Extended Data Fig. 6f). We then examined the transcription strengths of the variants that contained a perfect match to the consensus, and saw a wide range that varied from highly active to inactive (Extended Data Fig. 6g). These findings indicate that a perfect match to the RGWYGT consensus does not accurately predict the strength of the DPR. By contrast, we compared the SVRb scores to the observed transcription strengths of the same variants and saw an excellent correlation ($\rho = 0.95$) (Extended Data Fig. 6h). Thus, an SVR model is more effective than a standard consensus approach for predicting the activity of a sequence motif.

We also compared SVRb scores to the HOMER motif scores, which are based on the position-weight matrix (PWM) associated with the top HOMER consensus sequence (Extended Data Fig. 6i). These results showed that the comprehensive computational SVR model ($\rho = 0.90$) more accurately describes the DPR than the traditional consensus-based method ($\rho = 0.51$). The effectiveness of the SVR approach may be due, at least in part, to the training of the SVR with the full range of DPR sequences (that is, strong, intermediate, and weak), which is in contrast to the use of only strong variants in the generation of a consensus sequence.

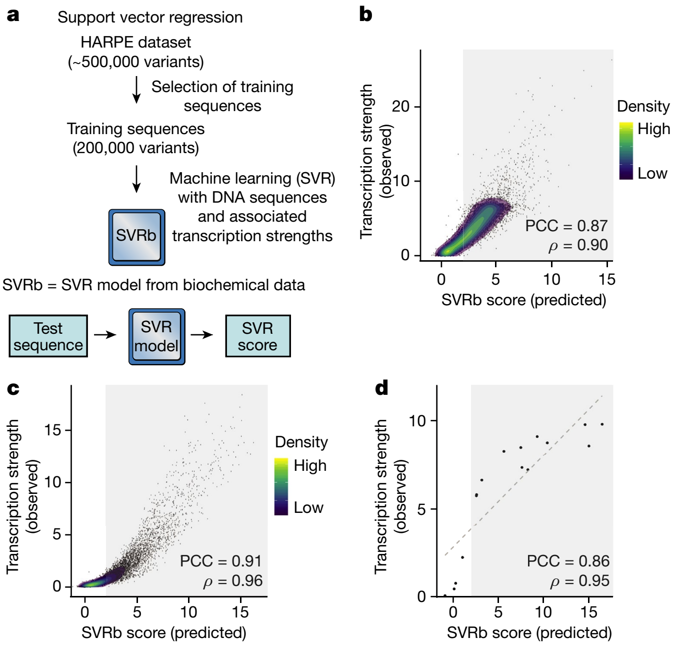


Fig. 3 | Machine learning analysis of the HARPE data yields an SVR model for the DPR. a, Summary of the SVR workflow. The HARPE dataset comprises about 500,000 DPR sequence variants, each with its associated transcription strength. A subset of these data (200,000 variants) was used to generate an SVR model for the DPR. The resulting SVR model was termed SVRb because it was trained with biochemical data. The SVR model provides a numerical score for the predicted transcription strength of any test sequence. **b**–**d**, To test the effectiveness of SVRb, the experimentally observed transcription strengths of sequence variants were compared with their predicted SVRb scores. **b**, Analysis of 7,500 independent test sequences in the HARPE dataset that were not used in the training of SVRb. The light grey shading (SVRb score ≥ 2) indicates predicted DPR activity (representative experiment, $n = 2$ biologically independent samples). **c**, Analysis of an independently generated HARPE dataset of a low-complexity DPR library (8,431 sequence variants) with high-confidence transcription strengths (representative experiment, $n = 2$ biologically independent samples). For **b**, **c**, PCC, Pearson's correlation coefficient with two-tailed $P < 2.2 \times 10^{-16}$; ρ, Spearman's rank correlation coefficient with two-tailed $P < 2.2 \times 10^{-16}$. **d**, Analysis of 16 DPR sequence variants (not in the training set) that were each tested individually by in vitro transcription and primer extension methodology (representative experiment, $n = 4$ biologically independent samples). PCC, Pearson's correlation coefficient with two-tailed $P = 3.4 \times 10^{-7}$; ρ, Spearman's rank correlation coefficient with two-tailed $P < 2.2 \times 10^{-16}$. For gel source data, see Supplementary Fig. 1.

Unlike a consensus-based model, the SVRb model can accurately incorporate the influence of neighbouring sequences on DPR activity (Extended Data Fig. 6j, k, Supplementary Discussion 2). We also found that SVR models can detect the function of an important sequence motif, such as a DPE-like sequence or a TATA motif, that is located at different positions within a larger region of interest (Extended Data Fig. 7a–i, Supplementary Discussion 3). In addition, SVRb uses information from a broader region of the DPR than a consensus-based model (Extended Data Fig. 7j, k, Supplementary Discussion 4). These findings thus indicate that SVR models are more effective at predicting transcription activity than consensus-based models.

SVR models from cell-based data

To test the versatility of SVR in the description of core promoter motifs, we compared SVR models created with HARPE data generated in vitro and in cells. With the DPR, we made SVRc (SVR of the DPR with cell-based data; the performance assessment of SVRc is in Extended Data

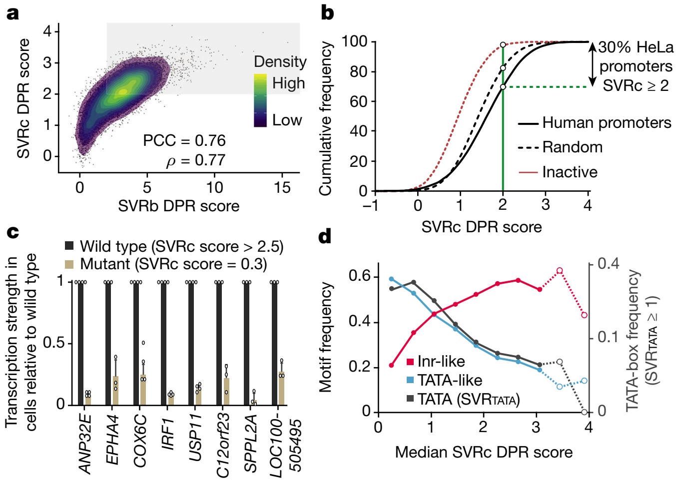


Fig. 4 | The DPR in human promoters. a, The SVR model from HARPE data in cells (SVRc) is similar to SVRb (biochemical). The SVRb and SVRc DPR scores of 7,500 test sequences (Fig. 3b) are compared. PCC, Pearson's correlation coefficient with two-tailed $P < 2.2 \times 10^{-16}$; ρ, Spearman's rank correlation coefficient with two-tailed $P < 2.2 \times 10^{-16}$. The light grey shading (SVRb and SVRc scores ≥ 2) indicates predicted DPR activity. **b**, Cumulative frequency of SVRc DPR scores in natural human promoters. Approximately 30% of 11,932 human promoters[33], 17% of 100,000 random sequences (61% average G/C content, as in human core promoters), and 2.6% of 10,000 inactive sequences (randomly selected from the 50% least active sequences in the HARPE assay) have an SVRc score of at least 2 (green line), which corresponds to an active DPR (Extended Data Fig. 5b). **c**, Mutational analysis reveals DPR activity in different human promoters (for genes shown on *x*-axis) with SVRc DPR scores >2.5. In the mutant promoters, the wild-type DPR was substituted with a DNA sequence that has an SVRc DPR score of 0.3 (Extended Data Fig. 4a). The promoter sequences are shown in Extended Data Fig. 8h. Promoter activity was measured by transient transfection in cells followed by primer extension analysis of the TSSs (data shown as mean ± s.d., $n = 3$ or 4 biologically independent samples, indicated by points representing independent samples). All $P < 0.05$ (two-tailed paired Student's *t*-test). For gel source data, see Supplementary Fig. 1. **d**, The SVRc DPR score correlates inversely with the presence of TATA-like sequences in human promoters in HeLa cells. The frequency of occurrence of Inr-like sequences, TATA-like sequences[12], and TATA-box motifs (SVRTATA ≥ 1) (Extended Data Fig. 5c) in human promoters that were binned according to their SVRc DPR scores (Extended Data Fig. 9a). Bins with fewer than 100 promoters are indicated with open circles and are connected by dashed lines (representative experiment, $n = 2$ biologically independent samples).

Fig. 5i–m), which correlated well ($\rho = 0.71$) with transcription strengths in cells and was reproducible ($\rho = 0.85$) (Extended Data Fig. 7l, m). Moreover, SVRc correlated well ($\rho = 0.77$) with SVRb in predicting the transcription strengths of DPR sequences (Fig. 4a).

With the TATA box, we used HARPE data generated in vitro and in cells (Extended Data Figs. 2k–p, 8a, b) to create SVR models (with the long TATA sequence) termed SVRTATA (in vitro) and SVRTATA (in cells) (Extended Data Fig. 7d–f; performance assessment of SVRTATA (in vitro) is shown in Extended Data Fig. 5n–r). SVRTATA (in vitro) was found to correlate well ($\rho = 0.86$) with transcription strengths as well as with SVRTATA (in cells) ($\rho = 0.80$) (Extended Data Fig. 7d, e). These results indicate that the use of HARPE in conjunction with SVR analysis is an effective method for the analysis of core promoter motifs. Furthermore, the extensive correlation between the in vitro and cell-based data (Figs. 2d, 4a, Extended Data Figs. 2g–j, 7d, 8a, b) provides comprehensive evidence that the mechanisms of transcription initiation in vitro are similar to those in cells.

The DPR is widely used in human promoters

To assess the role of the DPR in humans, we examined the relation between the HARPE-based DPR data and the corresponding sequences in natural human core promoters. First, we found that the relative nucleotide preferences in focused human core promoters[12] are similar to those in the most active sequences in the HARPE assay in vitro and in cells (Extended Data Fig. 8c–e). It is therefore likely that data from the HARPE assay reflect the properties of the DPR in natural human promoters.

By using the SVR models, we were able to estimate the occurrence of core promoter motifs in natural human focused promoters. With SVR models for the DPR, we found that about 25–34% of human promoters in different cell lines (HeLa, MCF7 and GM12878) are predicted to have an active DPR (Fig. 4b, Extended Data Fig. 8f, g, Supplementary Discussion 5). Similarly, with the SVRTATA models, we determined that about 15–23% of human promoters contain an active TATA box (Extended Data Fig. 7g–i, Supplementary Discussion 5). Thus, the DPR appears to be a widely used core promoter element. Moreover, the estimated occurrence of the DPR is comparable to that of the TATA box.

Notably, in sharp contrast to the DPR, a correctly positioned match to the RGWYGT DPE-like sequence[14] (Fig. 1c) was found in only about 0.4–0.5% of human focused promoters (Supplementary Discussion 5). Therefore, in humans, a consensus DPE-like sequence is rare, as previously noted[1–3], but the SVR-based DPR is relatively common. These findings further highlight the utility of machine learning relative to consensus approaches for the identification of core promoter sequence motifs.

We also tested the activities of individual DPR-like sequences in natural human promoters. To this end, we identified eight human promoters with an SVRc score of at least 2.5 and determined the activities of wild-type and mutant versions of the core promoters in cells (Fig. 4c, Extended Data Fig. 8h) and in vitro (Extended Data Fig. 8h, i). In all of the promoters that were tested, mutation of the DPR region resulted in a substantial decrease in transcriptional activity. These findings show that functionally active DPR motifs can be identified in natural promoters by using the SVR models.

Duality between the DPR and TATA box

To investigate the relation between the DPR, the TATA box, and the Inr, we examined the co-occurrence of these motifs in human promoters (Fig. 4d, Extended Data Fig. 9, Supplementary Discussion 6). We typically observed an increase in the occurrence of the Inr and Inr-like sequences with an increase in the SVR scores for the DPR. This effect is consistent with the cooperative function of the DPE and Inr motifs in *Drosophila*[13]. By contrast, the TATA motif is enriched in promoters lacking a DPR and depleted in promoters with high DPR scores. Similarly, but to a lesser extent, strong DPR motifs are more abundant in TATA-less promoters than in TATA-containing promoters (Extended Data Fig. 10). These findings suggest that some human core promoters depend predominantly on the DPR, whereas others depend mostly on the TATA box. This duality between the human DPR and TATA box suggests that they might have different biological functions and is consistent with the mutually exclusive properties of the DPE and TATA box in *Drosophila*[7,29–32]. Hence, the TATA–DPR duality is likely to reflect different mechanisms of transcription and potentially different modes of regulation of TATA-dependent versus DPR-dependent promoters in humans.

Here, we have used machine learning to decipher a promoter motif that could not be identified by the analysis of overrepresented sequences (Supplementary Discussion 7). Beyond the study of core promoters, this work describes a strategy for the machine learning analysis of functionally important DNA sequence motifs. In the future, it seems likely that machine learning models will continue to supersede consensus sequences in the characterization of DNA sequence motifs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2689-7.

1. Sandelin, A. et al. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet*. **8**, 424–436 (2007).
2. Vo ngoc, L., Wang, Y.-L., Kassavetis, G. A. & Kadonaga, J. T. The punctilious RNA polymerase II core promoter. *Genes Dev*. **31**, 1289–1301 (2017).
3. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol*. **19**, 621–637 (2018).
4. Meylan, P., Dreos, R., Ambrosini, G., Groux, R. & Bucher, P. EPD in 2020: enhanced data visualization and extension to ncRNA promoters. *Nucleic Acids Res*. **48** (D1), D65–D69 (2020).
5. Roeder, R. G. 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. *Nat. Struct. Mol. Biol*. **26**, 783–791 (2019).
6. Butler, J. E. & Kadonaga, J. T. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev*. **15**, 2515–2519 (2001).
7. Juven-Gershon, T., Hsu, J. Y. & Kadonaga, J. T. Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. *Genes Dev*. **22**, 2823–2830 (2008).
8. Zabidi, M. A. et al. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2015).
9. Parry, T. J. et al. The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. *Genes Dev*. **24**, 2013–2018 (2010).
10. Wang, Y. L. et al. TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes Dev*. **28**, 1550–1555 (2014).
11. Duttke, S. H. C., Doolittle, R. F., Wang, Y.-L. & Kadonaga, J. T. TRF2 and the evolution of the bilateria. *Genes Dev*. **28**, 2071–2076 (2014).
12. Vo Ngoc, L., Cassidy, C. J., Huang, C. Y., Duttke, S. H. & Kadonaga, J. T. The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. *Genes Dev*. **31**, 6–11 (2017).
13. Burke, T. W. & Kadonaga, J. T. *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev*. **10**, 711–724 (1996).
14. Kutach, A. K. & Kadonaga, J. T. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol. Cell. Biol*. **20**, 4754–4764 (2000).
15. Lim, C. Y. et al. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev*. **18**, 1606–1617 (2004).
16. Theisen, J. W. M., Lim, C. Y. & Kadonaga, J. T. Three key subregions contribute to the function of the downstream RNA polymerase II core promoter. *Mol. Cell. Biol*. **30**, 3471–3479 (2010).
17. Burke, T. W. & Kadonaga, J. T. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev*. **11**, 3020–3031 (1997).
18. Louder, R. K. et al. Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604–609 (2016).
19. Patel, A. B. et al. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362**, eaau8872 (2018).
20. Patwardhan, R. P. et al. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat. Biotechnol*. **27**, 1173–1175 (2009).
21. Lubliner, S. et al. Core promoter sequence in yeast is a major determinant of expression level. *Genome Res*. **25**, 1008–1017 (2015).
22. Arnold, C. D. et al. Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. *Nat. Biotechnol*. **35**, 136–144 (2017).
23. van Arensbergen, J. et al. Genome-wide mapping of autonomous promoter activity in human cells. *Nat. Biotechnol*. **35**, 145–153 (2017).
24. Weingarten-Gabbay, S. et al. Systematic interrogation of human promoters. *Genome Res*. **29**, 171–183 (2019).
25. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
26. Juven-Gershon, T., Cheng, S. & Kadonaga, J. T. Rational design of a super core promoter that enhances gene expression. *Nat. Methods* **3**, 917–922 (2006).
27. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn*. **20**, 273–297 (1995).
28. Vapnik, V. N. *The Nature of Statistical Learning Theory* (Springer, 1995).
29. Willy, P. J., Kobayashi, R. & Kadonaga, J. T. A basal transcription factor that activates or represses transcription. *Science* **290**, 982–985 (2000).
30. Hsu, J. Y. et al. TBP, Mot1, and NC2 establish a regulatory circuit that controls DPE-dependent versus TATA-dependent transcription. *Genes Dev*. **22**, 2353–2358 (2008).
31. Chen, K. et al. A global change in RNA polymerase II pausing during the *Drosophila* midblastula transition. *eLife* **2**, e00861 (2013).
32. Kedmi, A. et al. *Drosophila* TRF2 is a preferential core promoter regulator. *Genes Dev*. **28**, 2163–2174 (2014).
33. Duttke, S. H. C. et al. Human promoters are intrinsically directional. *Mol. Cell* **57**, 674–684 (2015).

Methods

HARPE screening vector and promoter inserts

The HARPE screening vector (Extended Data Fig. 1b) was created by modification of the SuRE plasmid[23] (a gift from J. van Arensbergen and B. van Steensel, Netherlands Cancer Institute). New features of the HARPE vector are as follows. First, to increase transcription levels, two GC-boxes (GGGGCGGGGC; binding sites for transcription factor Sp1) are located at positions −80 and −51 (the numbers indicate the positions of the upstream G of each GC-box) relative to the A_{+1} in the initiator (Inr) sequence of the core promoter that is to be inserted into the vector. Second, a TATA-like sequence (TTAACTATAA) upstream of the GC-boxes was mutated to CTGACTGGAC. Third, a KpnI restriction site is downstream of the −51 GC-box. Fourth, the KpnI site is followed by a spacer sequence and an AatII restriction site for insertion of core promoter sequences between the KpnI and AatII sites. Fifth, downstream of the AatII site, there is an RNA polymerase III (Pol III) terminator sequence (TTTTTTT) upstream of the transcribed sequence that is complementary to the reverse transcription primer. The Pol III terminator minimizes any potential background signal from Pol III transcription. For HARPE screening of randomized upstream sequences such as the TATA box, we used a slightly different screening vector in which the KpnI site is upstream of position −51. In this case, the downstream GC-box is included in the promoter insert rather than in the vector.

Randomized promoter inserts were generated by 5′ phosphorylation (T4 polynucleotide kinase; New England Biolabs) and annealing of partially complementary oligonucleotides (Extended Data Fig. 1c). The double-stranded DNA products were designed with 3′-overhangs for insertion between the KpnI and AatII sites of the HARPE vector. The SCP1m and human IRF1 core promoter sequences that were used are shown in Supplementary Table 2. In the analysis of the DPE region, the SCP1m region between +18 and +22 (CGAGC) was mutated to ATCCA (mutant MTE[26]). In the analysis of the TATA region, the SCP1m region between +28 and +34 (AGACGTG) was mutated to CTCATGT (mutant DPE[6]). In the IRF1 sequence, we introduced an A_{+11} to T substitution to eliminate a partial Pol III box A-like sequence.

HARPE library generation

The methodology for the preparation of the HARPE library was adapted from the SuRE procedure[23]. Annealed and phosphorylated promoter inserts were ligated into KpnI- and AatII-digested HARPE vector by using the TAKARA DNA Ligation Kit, Version 1 (Takara Bio). The resulting DNA was electroporated into DH5G CloneCatcher Gold (Genlantis) bacteria as recommended by the manufacturer, and the number of transformants was assessed by plating. Typically, a complexity of about 1,000,000 to 80,000,000 transformants was achieved. Next, a secondary downscaling step was performed to decrease the complexity of the library to about 100,000 or about 500,000 for shorter (8 to 12 bp) or longer (19 bp) randomized regions, respectively. Isolation of the DNA yielded the final HARPE DNA libraries, which were then transcribed in HeLa cells or in vitro.

Transcription of HARPE libraries in cells

HeLa cells (kind gift from the laboratory of A. Rao, La Jolla Institute for Immunology) were maintained at 37 °C under 5% CO_2 in DMEM (Gibco) supplemented with 10% FBS (ATCC), 50 U/ml penicillin (Thermo Fisher Scientific), and 50 μg/ml streptomycin (Thermo Fisher Scientific). HeLa cells were not authenticated but were tested and found to be negative for mycoplasma contamination. Transfections were performed with Lipofectamine 3000 (Thermo Fisher Scientific) as recommended by the manufacturer. Typically, two 10-cm culture dishes were used per sample. During collection, one-third of the cell pellet was reserved for plasmid DNA extraction, whereas the rest of the cells were used for RNA extraction. RNA processing was then performed as described below. All HARPE experiments in cells were performed independently two times to ensure reproducibility of the data. Replicates originated from the same HARPE DNA libraries that underwent independent transfection and downstream processing.

Transcription of HARPE libraries in vitro

For each sample library, the products from 12 standard in vitro transcription reactions were combined. Standard reactions were performed as follows. DNA template (500 ng) was incubated with HeLa nuclear extract[34] for preinitiation complex assembly at 30 °C for 1 h in 46 μl transcription buffer (20 mM HEPES-K^+ (pH 7.6), 50 mM KCl, 6 mM $MgCl_2$, 1.25% (w/v) polyvinyl alcohol, 1.25% (w/v) polyethylene glycol, 0.5 mM DTT, 3 mM ATP, 0.02 mM EDTA, and 2% (v/v) glycerol). rNTPs (4 μl; 0.4 mM final concentration of each rNTP) were added to initiate transcription. (Where indicated, sarkosyl was added to 0.2% (w/v) final concentration at 20 s after the addition of rNTPs.) The reaction was incubated at 30 °C for 20 min and terminated by the addition of 150 μl Stop Mix (20 mM EDTA, 200 mM NaCl, 1% (w/v) SDS, 0.3 mg/ml glycogen). Proteinase K (5 μl; 2.5 mg/ml) was added, and the mixture was incubated at 30 °C for 15 min. All in vitro transcription HARPE experiments were performed independently at least two times to ensure reproducibility of the data. Replicates originated from the same HARPE DNA libraries that underwent independent transcription and downstream processing.

RNA extraction and processing after transcription of HARPE libraries

RNA transcripts from cells or from in vitro transcription reactions were extracted with Trizol or Trizol LS (Thermo Fisher Scientific), respectively. Total RNA (40 μg for cell transfection experiments or the entire yield for in vitro experiments) was processed as follows. Contaminating plasmid DNA was removed with the TURBO DNA-free Kit—rigorous DNase treatment protocol (Thermo Fisher Scientific) as recommended by the manufacturer. The nucleic acids were precipitated with ethanol, and reverse transcription was performed with SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) with the RT primer (5′-GTGACTGGAGTTCAGACGTGT; Supplementary Table 2) as recommended by the manufacturer. The reaction products were then treated with 30 U RNase H (New England Biolabs) for 20 min at 37 °C. The nucleic acids were extracted with phenol-chloroform-isoamyl alcohol and precipitated with ethanol. The resulting cDNAs were then size-selected on a 6% polyacrylamide-8M urea gel using radiolabelled size markers (Supplementary Table 2) that enable the purification of cDNAs corresponding to transcription that initiates in the region from −5 to +6 relative to the A_{+1} in the Inr sequence.

Size-selected cDNAs were used as templates to generate DNA amplicons for Illumina sequencing using custom forward oligonucleotides containing the Illumina P5 and Read1-primer sequences preceding the sequence corresponding to nucleotides +1 to +16 of the promoter analysed (Supplementary Table 2). Reverse primers were selected from the NEBNext Multiplex Oligos for Illumina kits (NEB). NGS PCR amplicons were then size-selected on native 6% polyacrylamide gels before Illumina sequencing.

Processing of plasmid DNA for Illumina sequencing

For in vitro experiments, the starting material used was the HARPE DNA libraries. For cell transfection experiments, post-transfection plasmid DNA extraction was performed as described[23]. In brief, cells were treated with trypsin, washed with PBS, and then incubated in 500 μl nuclear extraction buffer (10 mM NaCl, 2 mM $MgCl_2$, 10 mM Tris-HCl (pH 7.8), 5 mM DTT, 0.5% NP40) on ice for 5 min. Nuclei were pelleted at 7,000*g* and washed twice with 1 ml nuclear extraction buffer. DNA was then extracted with ZymoPURE Plasmid Miniprep Kit (Zymo Research). Plasmid DNA samples were used as a template for the generation of DNA amplicons for Illumina sequencing. The forward oligonucleotides contain the Illumina P5 and Read1-primer sequences followed by a promoter-specific sequence (Supplementary Table 2) that comprises nucleotides +1 through +16 (relative to the +1 TSS) for accurate DNA count assessment. Reverse primers were selected from the NEBNext Multiplex Oligos for Illumina kits (New England Biolabs), which match the Illumina Read2-primer sequence present on the HARPE plasmid. NGS PCR amplicons were then size-selected on native 6% polyacrylamide gels before Illumina sequencing.

Illumina sequencing

Illumina sequencing of NGS PCR amplicons was carried out on a HiSeq 4000 or Novaseq 6000 at the IGM Genomics Center, University of California, San Diego, La Jolla, CA (Moores Cancer Center, supported by NIH grant P30 CA023100 and NIH SIG grant S10 OD026929).

Transcription of individual test sequences and candidate human promoters

The plasmids used for testing individual clones were constructed with the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) as recommended by the manufacturer. These constructs include core promoter sequences[12] from −36 to +50 nt relative to the +1 TSS of the specified genes.

For testing transcription activity in vitro, nucleic acids resulting from single standard reactions were isolated by phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation, and subjected to primer extension analysis with 5′-^{32}P-labelled RT primer. For testing transcription activity in cells, HeLa cells were transfected, and RNA was extracted with Trizol (Thermo Fisher Scientific). Total RNA (15 μg) was subjected to primer extension analysis with 5′-^{32}P-labelled RT primer.

Primer extension products were resolved on 6% polyacrylamide-8M urea gels and quantified by using a Typhoon imager (GE Health Sciences) and the associated Amersham Typhoon control software v1.1. Quantification of radiolabelled samples was measured with Fiji v1.52i. All experiments for individual clones were performed independently at least three times to ensure reproducibility of the data.

NGS data processing

Single-read sequences (SR75) were screened according to the following criteria: a perfect match to the 10 nt directly upstream of the randomized region followed by the exact nucleotide count within the randomized region and a perfect match to the 10 nt directly downstream of the randomized region. (For the analysis of the TATA box (long version), the SR75 sequencing reads only allowed for 8 nt following the barcode; thus, the criteria that we employed were as follows: perfect match to the 12 nt directly upstream of the barcode; exact size of randomized barcode; and perfect match to the 8 nt directly downstream of the barcode.) All reads containing a match to the selection pattern were deemed usable and trimmed for sequences outside the randomized region. When present, highly abundant reads in the randomized box that correspond to the original promoter sequence or to invariant sequences from other constructs were discarded, as they are likely to have originated from inaccurate indexing of other multiplexed samples. Read counts for each variant were then computed and yielded a plasmid DNA dataset (DNA dataset) and a cDNA dataset (RNA dataset) for each sample.

For each DNA dataset, we used only sequences with a minimum read count of 10 and a minimum relative count of 0.75 reads per million (RPM) so that low-confidence variants would not be included in the analysis. RNA dataset sequences were then matched to the corresponding DNA dataset, which was used as a reference. For each HARPE experiment, transcription strength was then defined as RNA tag count (in RPMs) divided by DNA tag count (in RPMs). Total read counts, number of variants, coverage values, and required DNA read counts are in Supplementary Table 1.

HARPE targeting the TATA box

HARPE libraries for the analysis of the TATA-box region were prepared using the same methodology as for the other HARPE libraries, except that a second randomized 'barcode' box was added between +53 and +63 nt (short TATA version) or +53 and +67 nt (long TATA version). The SCP1m region between +28 and +34 nt (AGACGTG) was also mutated to CTCATGT (mutant DPE[13]). Conversion tables from barcode to TATA-box variant were built by paired-end sequencing of amplicons from the starting plasmid libraries. Sequencing reads were screened as described above and clusters for which both read 1 and read 2 passed the screening criteria were used to compute read counts. A minimum read count threshold was set so that ≥98% of barcodes were associated with a single TATA-box variant. Pairs that did not reach the threshold and the remaining 2% of unassigned barcodes were discarded. DNA datasets and RNA datasets for all TATA-box HARPE experiments were matched to their corresponding barcode-to-TATA conversion tables. All non-matching barcodes were not included. TATA variants associated with multiple barcodes were combined, and their transcription strengths were computed as the average transcription strength across the multiple barcodes.

Low-complexity, high-confidence HARPE dataset

Low-complexity libraries were generated by limiting the randomization of the DPR (that is, setting nucleotides +17 to +35 to TCGKYYKSSYWKKRMRTGC, which yields a maximum complexity of 8,192) as well as by adding a randomized 3-nt tag from +55 to +57 nt. The final library contained about 130,000 DPR-tag pairs, which resulted in a median value of 13 out of 64 possible 3-nt tags per DPR variant. The transcription strength for each DPR variant was computed by determining the average of the RNA tag count/DNA tag count values for all of the DPR–tag pairs for that variant.

Motif discovery

Motif discovery was performed using HOMER[25]. findMotifs.pl was used to search the 0.1% most transcribed HARPE sequences in the region of interest. Variants randomly selected from all tested sequences were used as background. We looked for 19-nt motifs in the DPR datasets and 12-nt motifs in the DPE only and MTE only datasets. Because the TATA box is not constrained to a single position, we did not specify a motif length for the TATA-box datasets. The homer2 find tool was used to retrieve the sequences matching the top motif as well as to compute position-weight-matrix-based HOMER motif scores. These sequences were then used to generate the sequence logo using WebLogo 3[35,36].

Data processing, statistics and graphical displays

All calculations (including Pearson's correlation coefficients, Spearman's rank correlation coefficients, *P* values, means, and standard deviations) were performed in the R environment (version 3.6.1) in Rstudio v1.1.463 with R packages ggplot2 v3.2.1, tidyr v1.0.0, dplyr v0.8.3 and rlist v0.4.6.1, or with Microsoft Excel. All replicate measurements were taken from distinct samples. Adobe Illustrator CS v11.0.0 was used to build figures.

Training of SVR models

Machine learning analyses were performed using functions of the R package e1071 (D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel and F. Leisch (2019). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (formerly: E1071), TU Wien. R package version 1.7-2. https://CRAN.R-project.org/package=e1071). For SVR training, we used the default radial basis function (RBF) kernel, which yielded the best results among those tested. Grid search was performed for hyperparameters *C* (cost) and gamma, and cross validation was done by using two independent sets of sequences that were not used for the training (Extended Data Fig. 3b–d). Nucleotide variables for HARPE variants were computed as four categories (A, C, G and T), known as factors in R. To build the SVR model, we used the nucleotide variables as the input features and transcription strength as the output variable. For SVRb (or SVRc), we set aside 7,500 (or 6,500) test sequences (with the full range of transcription strengths) and trained the SVR with 200,000 of the remaining sequences (Extended Data Fig. 3a). For SVRTATA, we set aside 5,000 test sequences (with the full range of transcription strengths) and trained the SVR with all remaining (232,713) sequence variants.

Use of the SVR models to predict transcription strength

The SVR models described in this study can be used to predict transcription strength with R by using the predict() function included in CRAN package e1071. Models are imported with readRDS(). Query sequence data must be formatted as follows. The variable names are V1 to V12 for SVRTATA (corresponding to positions −32 to −21) and V1 to V19 for SVRc and SVRb (corresponding to positions from +17 to +35). Query sequences are split with one nucleotide per column and one sequence per row. Each column must have at least one A, one C, one G and one T to ensure that all variables are read as four categories (A, C, G, T). Prediction using an SVR model and a query sequence will return an output 'SVR score' that is related to the transcription strength and set on an arbitrary scale.

To streamline use of the models, we also provide an R script named SVRpredict.R (requires R with CRAN packages e1071 and docopt). SVRpredict.R inputs a model file as well as a sequence file (12- or 19-letter words/sequences, one sequence per line), and outputs a new file with each sequence and its associated predicted transcription strength in an added column (SVR_score).

Position index

To assess the effect of each sequence position on the SVR score, we used the position index (Extended Data Fig. 7j, k), which is the maximal SVR score increase that can be attained by a single nucleotide substitution at each position of the DPR. Because the positional contribution is affected by the sequence context (that is, the nucleotides at other positions within the DPR), the average positional contribution in 200 DPR contexts (that is, sequences in 200 different natural human promoters) was used to determine the position index.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The HARPE data are available from Gene Expression Omnibus (GEO; accession number, GSE139635). We obtained 5′-GRO-seq files (GSE63872[33] and GSE90035[12]) and GRO-cap files (GSM1480321)[37] from the Gene Expression Omnibus website (https://www.ncbi.nlm.nih.gov/geo/). Source data are provided with this paper.

Code availability

All computational analyses were performed by using R version 3.6.1 and previously described packages, as noted in the Methods.

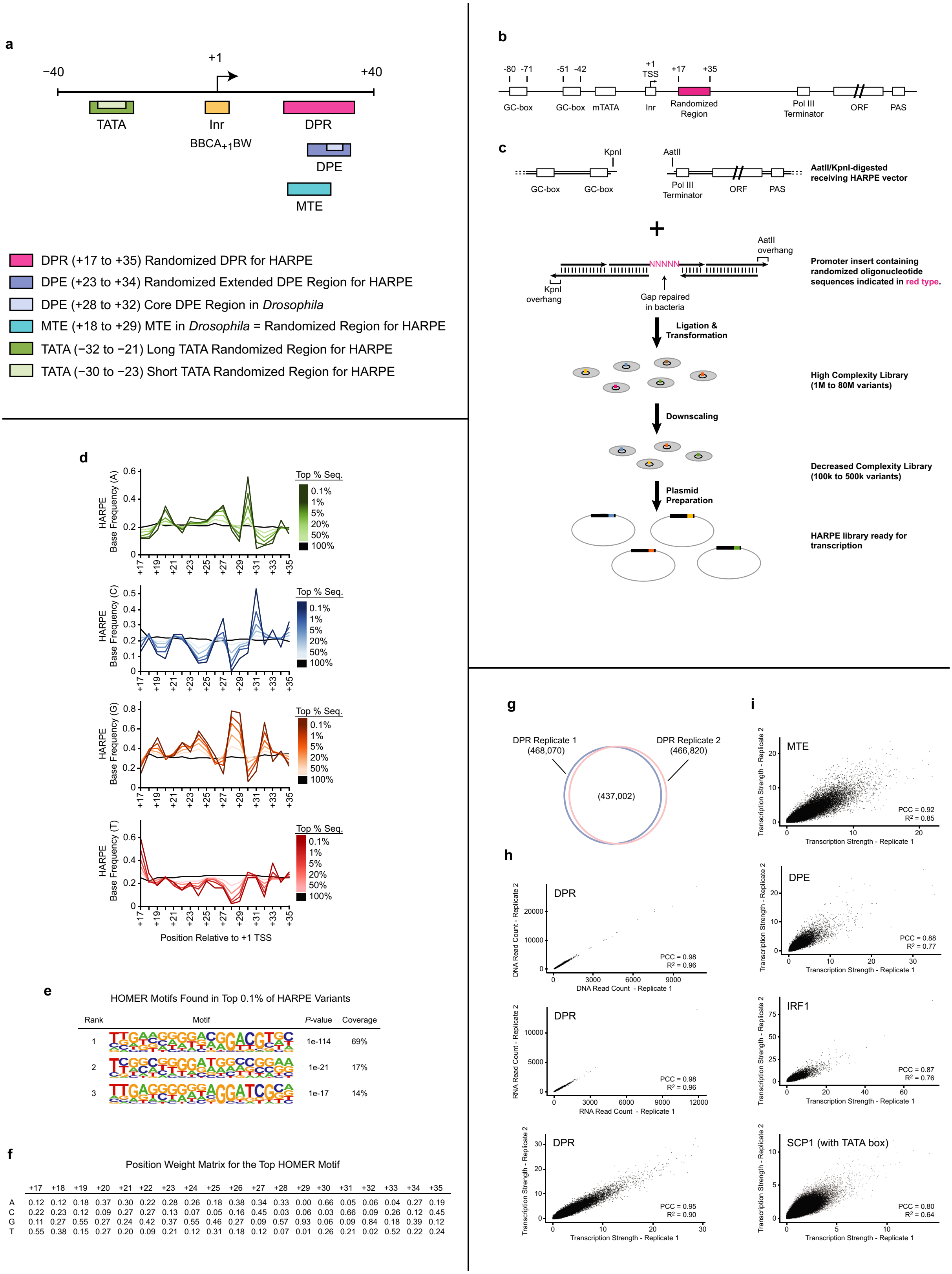
34. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res*. **11**, 1475–1489 (1983).
35. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res*. **14**, 1188–1190 (2004).
36. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*. **18**, 6097–6100 (1990).
37. Core, L. J. et al. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet*. **46**, 1311–1320 (2014).

Acknowledgements We thank E. P. Geiduschek, T. Juven-Gershon, G. Kassavetis, B. Delatte, J. Fei, G. Cruz-Becerra, and S. Chen for critical reading of the manuscript; J. van Arensbergen and B. van Steensel for the SuRE plasmid and protocols; B. Grant and C. Benner for advice; A. Rao for the HeLa cells; and the DNA sequencing facility at the Moores Cancer Center at UCSD (supported by NIH grant P30 CA023100 and NIH SIG grant S10 OD026929). L.V.n. received a UCSD Molecular Biology Cancer Fellowship. J.T.K. is the Amylin Chair in the Life Sciences. This work was supported by funding from NIH/NIGMS (R35 GM118060) to J.T.K.

Author contributions L.V.n., C.Y.H. and J.T.K. oversaw the overall design and execution of the project. The experiments were performed mostly by L.V.n. and C.Y.H. The analysis of the natural promoters was carried out by C.M. The computational analyses were performed by L.V.n., C.J.C. and C.Y.H. L.V.n. and J.T.K. were primarily responsible for writing the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2689-7.
Correspondence and requests for materials should be addressed to J.T.K.
Reprints and permissions information is available at http://www.nature.com/reprints.



e

HOMER Motifs Found in Top 0.1% of HARPE Variants

Rank	Motif	*P*-value	Coverage
1		1e-114	69%
2		1e-21	17%
3		1e-17	14%

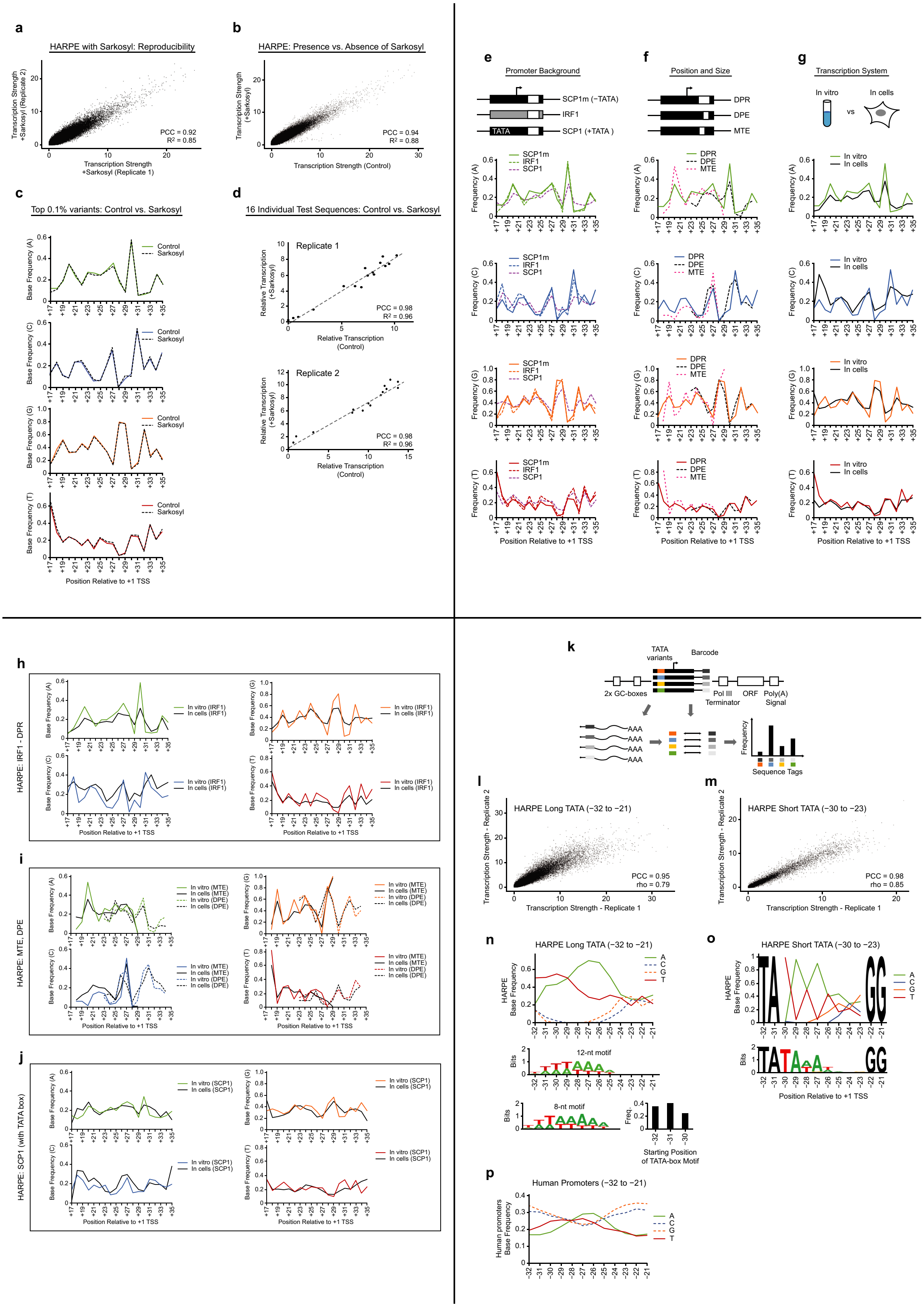
f

Position Weight Matrix for the Top HOMER Motif

	+17	+18	+19	+20	+21	+22	+23	+24	+25	+26	+27	+28	+29	+30	+31	+32	+33	+34	+35
A	0.12	0.12	0.18	0.37	0.30	0.22	0.28	0.26	0.18	0.38	0.34	0.33	0.00	0.66	0.05	0.06	0.04	0.27	0.19
C	0.22	0.23	0.12	0.09	0.27	0.27	0.13	0.07	0.05	0.16	0.45	0.03	0.06	0.03	0.66	0.09	0.26	0.12	0.45
G	0.11	0.27	0.55	0.27	0.24	0.42	0.37	0.55	0.46	0.27	0.09	0.57	0.93	0.06	0.09	0.84	0.18	0.39	0.12
T	0.55	0.38	0.15	0.27	0.20	0.09	0.21	0.12	0.31	0.18	0.12	0.07	0.01	0.26	0.21	0.02	0.52	0.22	0.24

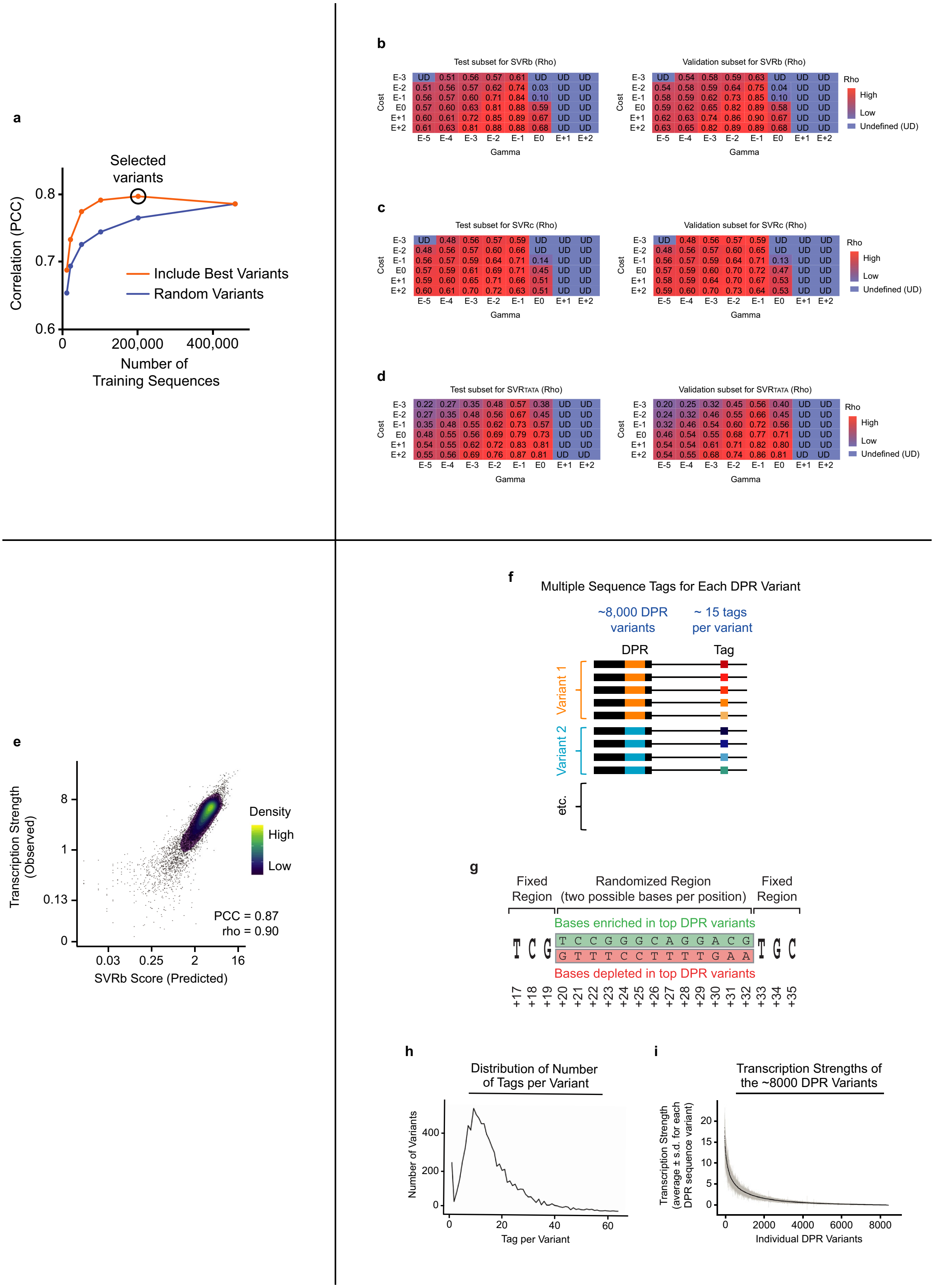
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Design and initial characterization of the HARPE assay. a, RNA polymerase II core promoter elements that were examined in this study. This diagram shows the positions of the TATA box, initiator (Inr), motif ten element (MTE), downstream core promoter element (DPE), and downstream core promoter region (DPR) relative to the A+1 nucleotide in the Inr consensus sequence. The Inr and MTE function together with a strict spacing requirement between the two motifs. The Inr and DPE similarly act together with a strict spacing requirement between the motifs. The Figure is drawn roughly to scale. The sequences that were randomized in the HARPE experiments are also indicated. **b**, **c**, Preparation of the HARPE library. **b**, HARPE constructs have two GC-boxes (Sp1 binding sites) upstream of the core promoter. The core promoters used in this study (SCP1m and IRF1) are TATA-less (mTATA = mutant TATA box), initiator (Inr)-containing promoters. An RNA polymerase III (Pol III) terminator prevents transcription by Pol III. The open reading frame of green fluorescent protein (ORF) and the polyadenylation signal (PAS) promote the synthesis of mature and stable transcripts. For the study of the DPR, the randomized region is from +17 to +35 relative to the +1 TSS. **c**, The fragments containing randomized elements are produced by annealing oligonucleotides that give protruding ends matching the KpnI and AatII sticky ends on the pre-digested plasmid. A high-complexity library of ~1M to 80M variants is typically obtained after bacterial transformation. If required, the level of complexity is decreased to ~100k to ~500k variants with a subset of the transformants. **d**, Nucleotide preferences can be observed in the most active DPR sequences. The nucleotide frequencies at each position of the DPR in the top 50% to the top 0.1% of the most transcribed sequences are indicated. All sequences (100%) are included as a reference. **e**, **f**, DPR motifs identified by HOMER. **e**, HOMER motifs found in the top 0.1% of HARPE DPR variants. **f**, Position-weight matrix for the top HOMER motif. P-values associated with hypergeometric tests (one tailed, no adjustment). All panels show a representative experiment (*n* = 2 biologically independent samples). **g**–**i**, HARPE is highly reproducible. **g**, Most variants are present and detectable in biological replicates. The intersection comprises variants detected in both biological replicates (exact sequence match). PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. **h**, Reproducibility of the DNA and RNA tag counts, and the resulting transcription strength value, for variants detected in both biological replicates. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. **i**, Reproducibility of the MTE, DPE, IRF1, and SCP1 (with TATA box) datasets, for variants detected in both biological replicates. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$.



Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Further characterization of the HARPE assay and modification of the HARPE assay to include the analysis of the upstream TATA box element. a–**d**, Relative promoter strengths in HARPE experiments performed in the absence versus the presence of sarkosyl. In vitro transcription reactions were performed in the absence or presence of 0.2% (w/v) sarkosyl (added immediately after transcription initiation). **a**, HARPE datasets with reactions performed in the presence of sarkosyl are reproducible. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2\times10^{-16}$. **b**, Relative promoter strength does not appear to be affected by the addition of sarkosyl. Comparison of HARPE data from reactions carried out in the absence (Control) or the presence of sarkosyl. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2\times10^{-16}$. **c**, The top 0.1% most highly transcribed promoter variants show similar nucleotide preferences in the absence (Control) or the presence of sarkosyl (representative experiment, $n=2$ biologically independent samples). **d**, The individual analysis of 16 independent promoter variants shows that the relative promoter strengths are approximately the same in the absence (Control) or the presence of sarkosyl. PCC, Pearson's correlation coefficient with two-tailed P-value $=7.1\times10^{-11}$ (replicate 1) or 1.7×10^{-11} (replicate 2). For gel source data, see Supplementary Fig. 1. **e**–**g**, HARPE yields consistent data under different conditions. The nucleotide frequencies of the top 0.1% most active sequences are shown. **e**, HARPE analysis (in vitro) of the DPR with three different promoter cassettes: SCP1 lacking a TATA box (SCP1m), the human IRF1 core promoter (IRF1), and SCP1 containing a TATA box (SCP1). **f**, HARPE of the DPR (+17 to +35), DPE (+23 to +34), and MTE (+18 to +29) motifs with the SCP1m promoter in vitro. **g**, HARPE of the DPR in the SCP1m promoter transcribed in vitro or in cells. All panels show a representative experiment, $n=2$ biologically independent samples. **h**–**j**, HARPE data generated in cells are similar to the corresponding in vitro data. **h**, The nucleotide frequencies of the top 0.1% most active DPR sequences obtained in cells are consistent with their in vitro counterparts. These HARPE experiments were performed with the human IRF1 core promoter. **i**, The nucleotide frequencies of the top 0.1% most active MTE and DPE sequences obtained in cells are consistent with their in vitro counterparts. These experiments examined either the MTE region or the DPE region in cells or in vitro. **j**, The nucleotide frequencies of the top 0.1% most active DPR sequences obtained in cells are consistent with their in vitro counterparts. These HARPE experiments were performed with the TATA-box-containing SCP1 core promoter. All panels show a representative experiment ($n=2$ biologically independent samples). **k**–**p**, HARPE can be used to analyse regions upstream of the TSS. **k**, Design of a HARPE experiment targeting the upstream TATA-box region. Sequencing of the DNA constructs provides a correspondence between each TATA-box variant and a downstream barcode. Analysis of the barcode sequence in each transcript thus identifies its associated TATA-box variant sequence. **l**, HARPE was performed with a randomized region from −32 to −21 (long TATA) relative to the +1 TSS. The reproducibility of two independent experiments is shown. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2\times10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2\times10^{-16}$. **m**, HARPE was carried with a randomized region from −30 to −23 (short TATA) with an upstream TA dinucleotide at positions −32 and −31. The upstream TA sequence directs the formation of the TATA box in a single phase. The reproducibility of two independent experiments is also shown. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2\times10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2\times10^{-16}$. **n**, The nucleotide frequencies and top 8-nt and 12-nt HOMER motifs for the top 0.1% most transcribed variants are shown for HARPE data with the long TATA (−32 to −21) randomized sequence. The upstream T of the 8-nt TATA box motif was found to be located at position −32, −31, or −30 (representative experiment, $n=2$ biologically independent samples). **o**, The nucleotide frequencies and top 8-nt HOMER motif for the top 0.1% most transcribed variants are shown for HARPE data with the short TATA (−30 to −23) randomized sequence. In the short TATA analysis, the upstream T of the TATA box is fixed at position −32, and thus, a distinct TATA-box sequence can be seen in the HOMER analysis (representative experiment, $n=2$ biologically independent samples). **p**, The nucleotide frequencies in natural human focused promoters[12] are similar to those in the long TATA dataset (**n**), particularly with the A and T nucleotides.



Test subset for SVRb (Rho)

Cost \ Gamma	E-5	E-4	E-3	E-2	E-1	E0	E+1	E+2
E-3	UD	0.51	0.56	0.57	0.61	UD	UD	UD
E-2	0.51	0.56	0.57	0.62	0.74	0.03	UD	UD
E-1	0.56	0.57	0.60	0.71	0.84	0.10	UD	UD
E0	0.57	0.60	0.63	0.81	0.88	0.59	UD	UD
E+1	0.60	0.61	0.72	0.85	0.89	0.67	UD	UD
E+2	0.61	0.63	0.81	0.88	0.88	0.68	UD	UD

Validation subset for SVRb (Rho)

Cost \ Gamma	E-5	E-4	E-3	E-2	E-1	E0	E+1	E+2
E-3	UD	0.54	0.58	0.59	0.63	UD	UD	UD
E-2	0.54	0.58	0.59	0.64	0.75	0.04	UD	UD
E-1	0.58	0.59	0.62	0.73	0.85	0.10	UD	UD
E0	0.59	0.62	0.65	0.82	0.89	0.58	UD	UD
E+1	0.62	0.63	0.74	0.86	0.90	0.67	UD	UD
E+2	0.63	0.65	0.82	0.89	0.89	0.68	UD	UD

Test subset for SVRc (Rho)

Cost \ Gamma	E-5	E-4	E-3	E-2	E-1	E0	E+1	E+2
E-3	UD	0.48	0.56	0.57	0.59	UD	UD	UD
E-2	0.48	0.56	0.57	0.60	0.66	UD	UD	UD
E-1	0.56	0.57	0.59	0.64	0.71	0.14	UD	UD
E0	0.57	0.59	0.61	0.69	0.71	0.45	UD	UD
E+1	0.59	0.60	0.65	0.71	0.66	0.51	UD	UD
E+2	0.60	0.61	0.70	0.72	0.63	0.51	UD	UD

Validation subset for SVRc (Rho)

Cost \ Gamma	E-5	E-4	E-3	E-2	E-1	E0	E+1	E+2
E-3	UD	0.48	0.56	0.57	0.59	UD	UD	UD
E-2	0.48	0.56	0.57	0.60	0.65	UD	UD	UD
E-1	0.56	0.57	0.59	0.64	0.71	0.13	UD	UD
E0	0.57	0.59	0.60	0.70	0.72	0.47	UD	UD
E+1	0.58	0.59	0.64	0.70	0.67	0.53	UD	UD
E+2	0.59	0.60	0.70	0.73	0.64	0.53	UD	UD

Test subset for SVRTATA (Rho)

Cost \ Gamma	E-5	E-4	E-3	E-2	E-1	E0	E+1	E+2
E-3	0.22	0.27	0.35	0.48	0.57	0.38	UD	UD
E-2	0.27	0.35	0.48	0.56	0.67	0.45	UD	UD
E-1	0.35	0.48	0.55	0.62	0.73	0.57	UD	UD
E0	0.48	0.55	0.56	0.69	0.79	0.73	UD	UD
E+1	0.54	0.55	0.62	0.72	0.83	0.81	UD	UD
E+2	0.55	0.56	0.69	0.76	0.87	0.81	UD	UD

Validation subset for SVRTATA (Rho)

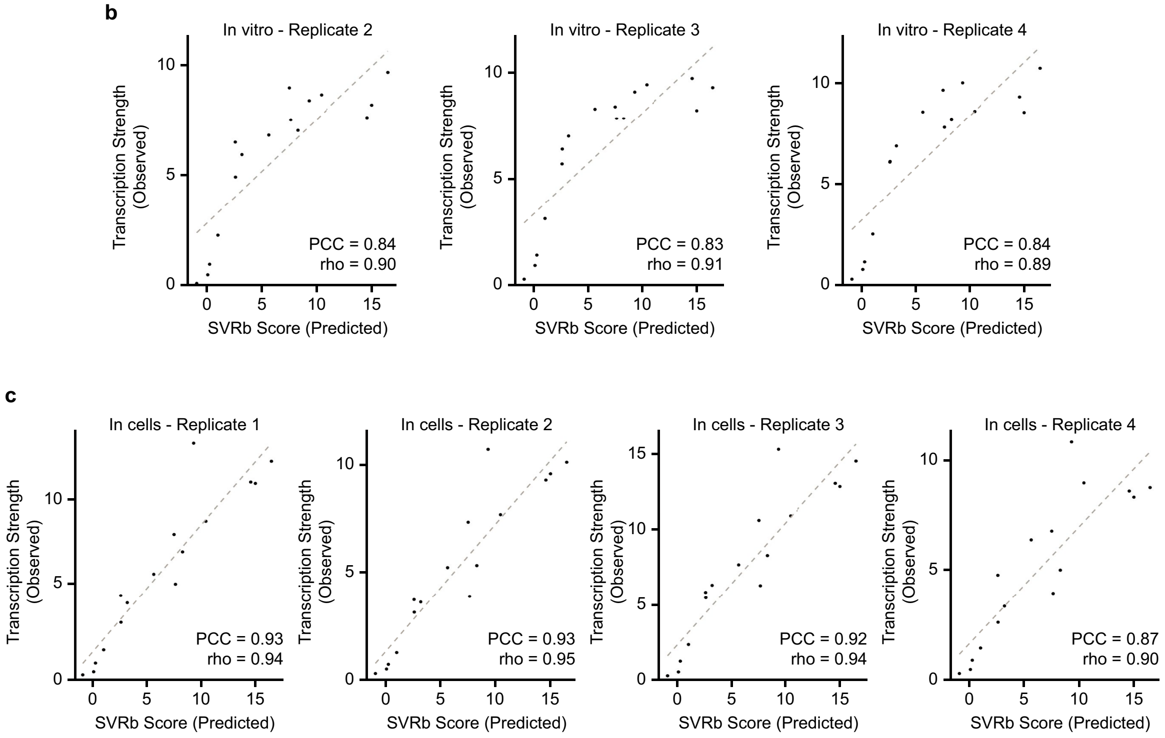
Cost \ Gamma	E-5	E-4	E-3	E-2	E-1	E0	E+1	E+2
E-3	0.20	0.25	0.32	0.45	0.56	0.40	UD	UD
E-2	0.24	0.32	0.46	0.55	0.66	0.45	UD	UD
E-1	0.32	0.46	0.54	0.60	0.72	0.56	UD	UD
E0	0.46	0.54	0.55	0.68	0.77	0.71	UD	UD
E+1	0.54	0.54	0.61	0.71	0.82	0.80	UD	UD
E+2	0.54	0.55	0.68	0.74	0.86	0.81	UD	UD

Extended Data Fig. 3 | See next page for caption.

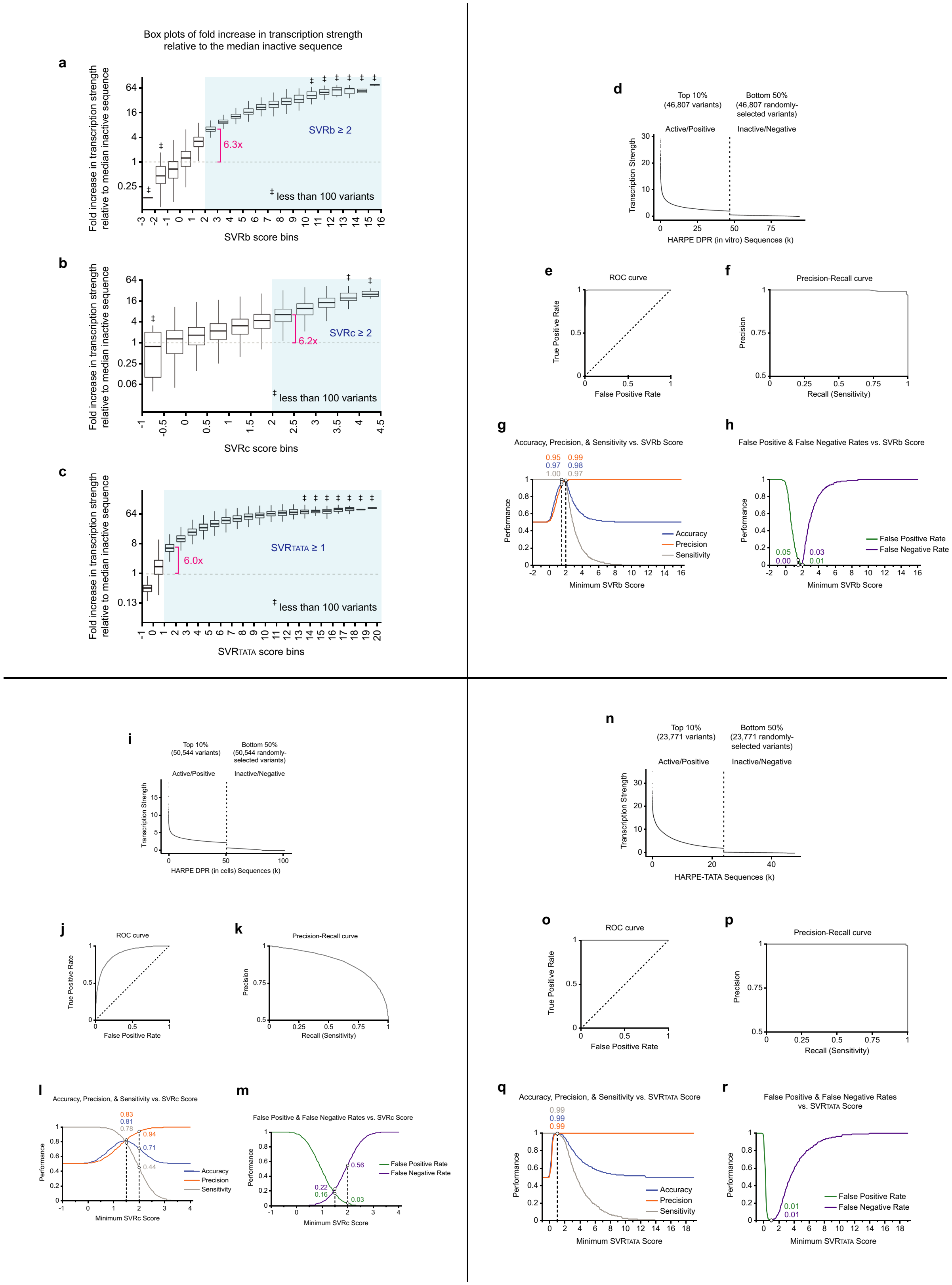
Extended Data Fig. 3 | Initial characterization and optimization of the SVR models and the creation of a low complexity HARPE library for further SVR analysis of the DPR. a, Selection of sequences for training of the SVR. Different numbers of training sequences were selected either randomly (blue line) or by using a combination of the most transcribed (Best) variants and Non-Best variants (that is, those variants that are not in the Best category) at a 1:1 ratio of Best:Non-Best (orange line). The resulting SVR models were used to predict the transcriptional activity of the Test Sequences in Fig. 3b, and the correlations between the predicted versus observed transcriptional activities are shown on the Y axis. In our studies, we used the SVR model (Selected variants) that was built on the training set that consists of the 100,000 most transcribed (Best) variants and randomly selected 100,000 Non-Best variants (representative experiment $n = 2$ biologically independent samples). The models in this figure were built by using default parameters for SVR training. **b**–**d**, Grid search cross validation for the SVR models. Grid search results with different values for the cost of misclassification (cost) and individual training example influence (gamma) for (**b**) SVRb, (**c**) SVRc, and (**d**) SVRTATA. Shown are Spearman's rank correlation coefficient (rho) between the prediction of each model and the observed transcription strength with two independent datasets (validation and test sets, which are separate halves of the test sequences described in Fig. 3b) that were not used in the training of the models. SVR models were trained as described in Methods. Undefined (UD) correlation is observed when the prediction of a model is constant regardless of the sequence. The hyperparameter values that were selected in this study are as follows: SVRb (c = 10 and gamma = 0.1); SVRc (c = 1, gamma = 0.02); and SVRTATA (c = 100, gamma = 0.1). **e**, Concordance between the predicted and observed activities of DPR sequence variants, as shown with a logarithmic scale. Analysis of 7500 independent test sequences in the HARPE dataset that were not used in the training of SVRb. This figure presents the data shown in Fig. 3b with a log scale for the *x*- and *y*-axes. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. **f**–**i**, Design and use of a low complexity HARPE library that provides high-quality data on 8,431 unique DPR variants. **f**, Design of a low complexity library with multiple DNA sequence tags for each DPR variant. A restricted library was built with 8,431 unique DPR variants. Each variant was associated with about 15 downstream DNA sequence tags that enable multiple measurements of transcription strength for the same variant within the same experiment. **g**, To restrict the complexity of the library, the randomized region was shortened to 13 nucleotides, and each position contained one of only two different bases. **h**, The number of tags per variant. The median value is 13 (representative experiment, $n = 2$ biologically independent samples). **i**, The observed transcription strength for each of the DPR variants. There are multiple different sequence tags for each DPR variant. The plot shows the average (black) ± standard deviation (designated in grey) for each of the variants (representative experiment, $n = 2$ biologically independent samples).

a

	+17	+18	+19	+20	+21	+22	+23	+24	+25	+26	+27	+28	+29	+30	+31	+32	+33	+34	+35	In vitro SVRb Score	In cells SVRc Score
SCP1m	**T**	**C**	**G**	**A**	**G**	**C**	**C**	**G**	**A**	**G**	**C**	**A**	**G**	**A**	**C**	**G**	**T**	**G**	**C**	**14.6**	**4.0**
	T	C	G	A	G	C	G	G	**G**	G	**A**	**G**	G	A	C	G	**C**	G	C	15.0	4.2
	T	C	G	A	G	C	C	G	A	G	**A**	**G**	G	A	C	G	T	G	C	16.5	4.1
	T	C	G	A	G	C	C	G	A	G	C	**C**	G	A	C	G	T	G	C	7.5	3.3
	T	C	G	A	G	C	C	G	A	G	**G**	A	G	A	C	G	T	G	C	9.3	3.1
	T	C	G	A	G	C	C	G	A	G	C	A	G	A	C	**T**	T	G	C	7.6	3.2
	T	C	G	A	G	C	C	G	A	G	C	A	G	**C**	C	G	T	G	C	8.3	3.1
	T	C	G	A	G	C	C	**T**	A	G	C	A	G	A	C	G	T	G	C	10.4	3.3
	T	C	G	A	G	C	C	G	A	G	C	A	**T**	A	C	G	T	G	C	5.7	2.8
	T	C	G	A	G	C	C	G	A	G	C	**C**	**T**	A	C	G	T	G	C	2.6	2.3
	T	C	G	A	G	C	C	G	A	G	C	A	**T**	A	C	**T**	T	G	C	3.2	2.1
	T	C	G	**T**	G	**T**	C	**T**	A	G	C	A	**T**	A	C	G	T	G	C	2.6	1.6
	T	C	G	A	G	C	C	**T**	A	G	C	**C**	**T**	A	C	**T**	T	G	C	1.1	1.1
	T	**A**	**T**	**C**	**T**	**T**	**T**	**T**	**C**	**T**	**G**	**C**	**T**	**C**	**G**	**T**	**A**	**C**	**A**	0.1	0.5
	T	**A**	**T**	A	G	C	C	**T**	A	G	**G**	**C**	**T**	**C**	C	**T**	T	G	C	0.3	0.3
	T	**A**	**A**	A	**A**	**A**	**A**	**A**	A	**A**	**A**	**T**	**A**	A	**A**	**T**	**A**	**A**	**A**	-0.9	-0.1

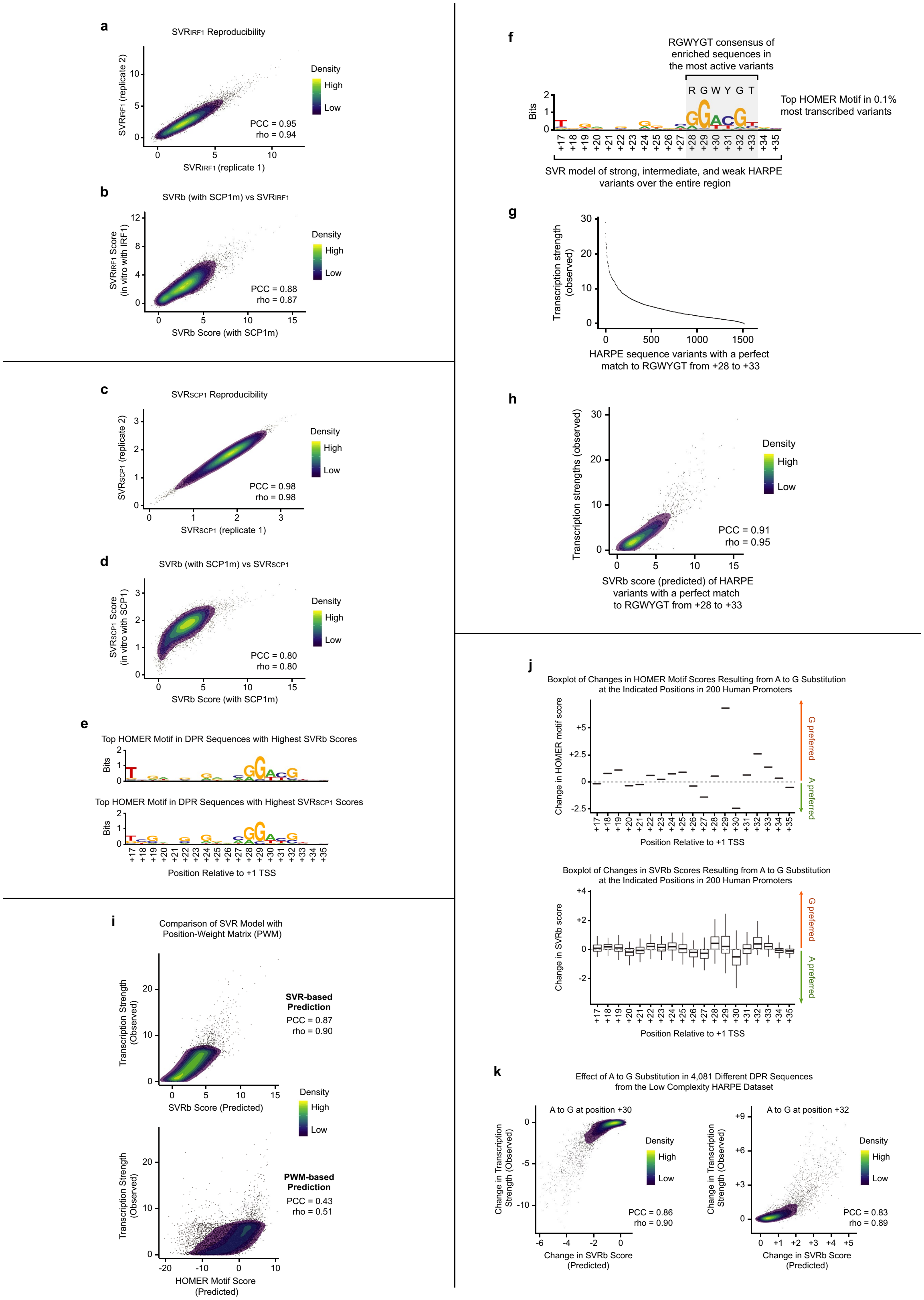


Extended Data Fig. 4 | Individual assessment of the transcription activity of 16 independent variants that are not present in the SVR training set. **a**, The 16 variants, which include the original SCP1m sequence, represent a wide range of SVR scores. Nucleotides that differ from the SCP1m sequence are indicated in red type. **b**, The 16 promoter sequences were inserted into plasmids and subjected to in vitro transcription and primer extension analysis (*n* = 4 biologically independent samples). The plots show the predicted SVRb scores and the observed transcription strengths. Replicate 1 is shown in Fig. 3d. PCC, Pearson's correlation coefficient with two-tailed P-values $<1.7 \times 10^{-6}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. For gel source data, see Supplementary Fig. 1. **c**, The 16 promoters were subjected to transient transfection and primer extension analysis (*n* = 4 biologically independent samples). The plots show the predicted SVRb scores and the observed transcription strengths. PCC, Pearson's correlation coefficient with two-tailed P-value $<3.9 \times 10^{-6}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. For gel source data, see Supplementary Fig. 1.



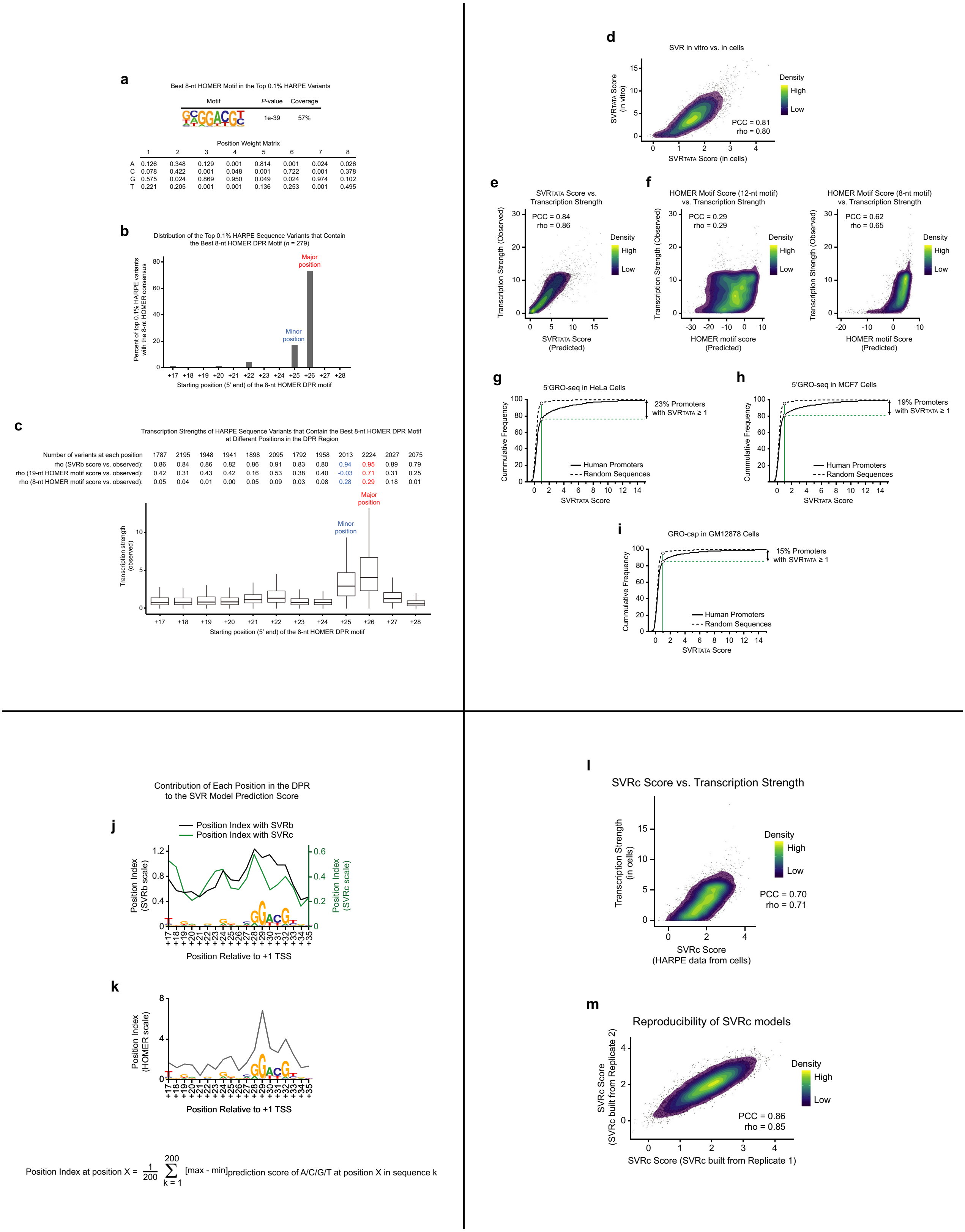
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Use of the SVR models to identify active sequence elements and performance assessment of the SVR models. a–**c**, The relationship between SVR scores and transcription strength. Box-plot diagrams are shown for (**a**) SVRb, (**b**) SVRc, and (**c**) SVRTATA with all of their corresponding HARPE sequence variants that are placed in bins of the indicated SVR score ranges. Sequence variants with SVRb score ≥ 2, SVRc score ≥ 2, and SVRTATA score ≥ 1 are typically at least about 6 times more active than an inactive sequence (light blue shaded regions), and are thus designated as "active". The thick horizontal lines are the medians, and the lower and upper hinges are the first and third quartiles, respectively. Each upper (or lower) whisker extends from the upper (or lower) hinge to the largest (or lowest) value no further than 1.5 * IQR from the hinge. Data beyond the end of the whiskers (outlying points) are omitted from the box plot. Sequence variants with transcription strength = 0 were removed to allow log-scale display of the diagrams. The horizontal dashed grey lines denote the transcription strengths of the median inactive sequences. **d**–**h**, Performance assessment of SVRb. All panels show a representative experiment ($n = 2$ biologically independent samples). **d**, Selection of HARPE variants used in performance assessment. The top 10% sequence variants were designated as active/positive for transcription, and an equal (randomly selected) number of the bottom 50% of sequence variants were designated as inactive/negative for transcription. These sequences were then used in the performance assessment. Intermediate variants that were between the top and bottom groups were not included. The transcription strengths of all selected sequences are shown. **e**, Receiver operating characteristic (ROC) curve. **f**, Precision-recall (PR) curve. **g**, Performance measures relative to the minimum SVRb score required for a positive prediction. Performance was computed by counting true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Accuracy [(TP+TN) / (TP+FP+TN+FN)] reflects how often SVRb predictions are correct. Precision [TP / (TP + FP)] is the proportion of positive predictions that are correct. Sensitivity or recall or true positive rate [TP / (TP + FN)] is the proportion of transcriptionally active variants that are correctly predicted as positives. **h**, False positive and false negative rates. The false positive rate [FP / (FP + TN)] is the probability for an inactive sequence to be incorrectly predicted as positive. The false negative rate [FN / (FN + TP)] = (1 − Sensitivity) is the probability for an active sequence to be incorrectly predicted as negative. Performance values are shown for selected minimum SVRb scores (1.5 and 2). All panels show a representative experiment ($n = 2$ biologically independent samples). **i**–**m**, Performance assessment of SVRc. **i**, Selection of HARPE variants used in performance assessment. The top 10% sequence variants were designated as active/positive for transcription, and an equal (randomly selected) number of the bottom 50% of sequence variants were designated as inactive/negative for transcription. These sequences were then used in the performance assessment. Intermediate variants that were between the top and bottom groups were not included. The transcription strengths of all selected sequences are shown. **j**, Receiver operating characteristic (ROC) curve. **k**, Precision-recall (PR) curve. **l**, Performance measures relative to the minimum SVRc score required for a positive prediction. Performance was computed by counting true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Accuracy [(TP+TN) / (TP+FP+TN+FN)] reflects how often SVRc predictions are correct. Precision [TP / (TP + FP)] is the proportion of positive predictions that are correct. Sensitivity [TP / (TP + FN)] is the proportion of transcriptionally active variants that are correctly predicted as positives. **m**, False positive and false negative rates. The false positive rate [FP / (FP + TN)] is the probability for an inactive sequence to be incorrectly predicted as positive. The false negative rate [FN / (FN + TP)] = (1 − Sensitivity) is the probability for an active sequence to be incorrectly predicted as negative. Performance values are shown for selected minimum SVRc scores (1.5 and 2). All panels show a representative experiment ($n = 2$ biologically independent samples). **n**–**r**, Performance assessment of SVRTATA. **n**, Selection of HARPE variants used in performance assessment. The top 10% sequence variants were designated as active/positive for transcription, and an equal (randomly selected) number of the bottom 50% of sequence variants were designated as inactive/negative for transcription. These sequences were then used in the performance assessment. Intermediate variants that were between the top and bottom groups were not included. The transcription strengths of all selected sequences are shown. One outlier variant with an exceptionally high transcription level was omitted in the graph, but was included in the performance analysis. **o**, Receiver operating characteristic (ROC) curve. **p**, Precision-recall (PR) curve. **q**, Performance measures relative to the minimum SVRTATA score required for a positive prediction. Performance was computed by counting true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Accuracy [(TP+TN) / (TP+FP+TN+FN)] reflects how often SVRTATA predictions are correct. Precision [TP / (TP + FP)] is the proportion of positive predictions that are correct. Sensitivity [TP / (TP + FN)] is the proportion of transcriptionally active variants that are correctly predicted as positives. **r**, False positive and false negative rates. The false positive rate [FP / (FP + TN)] is the probability for an inactive sequence to be incorrectly predicted as positive. The false negative rate [FN / (FN + TP)] = (1 − Sensitivity) is the probability for an active sequence to be incorrectly predicted as negative. Performance values are shown for minimum SVRTATA scores = 1.0. All panels show a representative experiment ($n = 2$ biologically independent samples).



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Further analysis of the SVR models and their relation to consensus sequence-based approaches. a–**e**, SVR models based on HARPE data with different promoter backgrounds are consistent. SVR models were tested with the 7500 DPR sequence variants used in Fig. 3b. **a**, SVRIRF1 models trained with HARPE data for the DPR with the IRF1 promoter cassette (promoter background) are reproducible. **b**, SVRb based on HARPE data for the DPR with the SCP1m promoter cassette (promoter background) is similar to the SVRIRF1 model trained with HARPE data for the DPR in the IRF1 background. **c**, SVRSCP1 models trained with HARPE data for the DPR with the SCP1 (TATA-containing) promoter cassette (promoter background) are reproducible. **d**, SVRb for the DPR in the TATA-less SCP1m promoter cassette (promoter background) is similar to the SVRSCP1 model for the DPR in the TATA-containing SCP1 promoter cassette. **e**, SVRb and SVRSCP1 exhibit similar DNA sequence preferences. This figure shows the web logos for the top HOMER motifs identified with the top 0.1% DPR sequences (in 500,000 random sequences), as assessed with either SVRb or SVRSCP1. **f**–**h**, SVR analysis incorporates information that is not encapsulated in a consensus of enriched sequences in the most active variants. **f**, Web logo for the top HOMER motif identified with the 0.1% most transcribed DPR sequences. This panel is adapted from Fig. 1c and shows the DPE-like RGWYGT consensus of enriched sequences from +28 to +33. In contrast, the SVR model is generated from strong, intermediate, and weak variants of the entire DPR region. **g**, HARPE variants with a perfect match to the RGWYGT consensus exhibit transcription strengths that range from highly active to inactive. **h**, SVRb accurately predicts the transcription strengths of different HARPE variants with a perfect match to the RGWYGT consensus. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. **i**, An SVR-based approach provides a more accurate prediction of DPR activity than a consensus sequence-based method. The plots show the correlation between the observed transcription strength (in vitro) and the predicted scores of the DPR, as assessed with either SVRb (upper; adapted from Fig. 3b) or a consensus sequence/position-weight matrix-based method (HOMER; lower). The HOMER consensus/position-weight matrix (Fig. 1c, Extended Data Fig. 1e, f) is based on the top 0.1% most transcribed DPR sequences. The DPR variants are the 7500 Test Sequences shown in Fig. 3. The coloured density scale is identical for both plots (representative experiment, *n* = 2 biologically independent samples). PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. **j**, **k**, SVRb scores are influenced by DNA sequence context (that is, flanking nucleotides), whereas PWM-based HOMER scores treat individual nucleotide positions independently. **j**, Box-plot diagrams of the changes in the HOMER motif scores (top) and the SVRb scores (bottom) due to an A-to-G substitution at each of the indicated positions. The values were generated with 200 different DPR sequences in randomly-selected natural human promoters. The thick horizontal lines are the medians, and the lower and upper hinges are the first and third quartiles, respectively. Each upper (or lower) whisker extends from the upper (or lower) hinge to the largest (or lowest) value no further than 1.5 * IQR from the hinge. Data beyond the end of the whiskers (outlying points) are omitted from the box plot. A representative experiment is shown (*n* = 2 biologically independent samples). **k**, The influence of sequence context is accurately captured by the SVR model. Shown are the changes in SVRb score and transcription strength for 4,081 DPR variants when A is mutated to G at positions +30 (left) or +32 (right). The transcription data of the sequence variants were from the Low Complexity Library (Fig. 3c). PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$.



Best 8-nt HOMER Motif in the Top 0.1% HARPE Variants

Motif	*P*-value	Coverage
GCGGACGT	1e-39	57%

Position Weight Matrix

	1	2	3	4	5	6	7	8
A	0.126	0.348	0.129	0.001	0.814	0.001	0.024	0.026
C	0.078	0.422	0.001	0.048	0.001	0.722	0.001	0.378
G	0.575	0.024	0.869	0.950	0.049	0.024	0.974	0.102
T	0.221	0.205	0.001	0.001	0.136	0.253	0.001	0.495

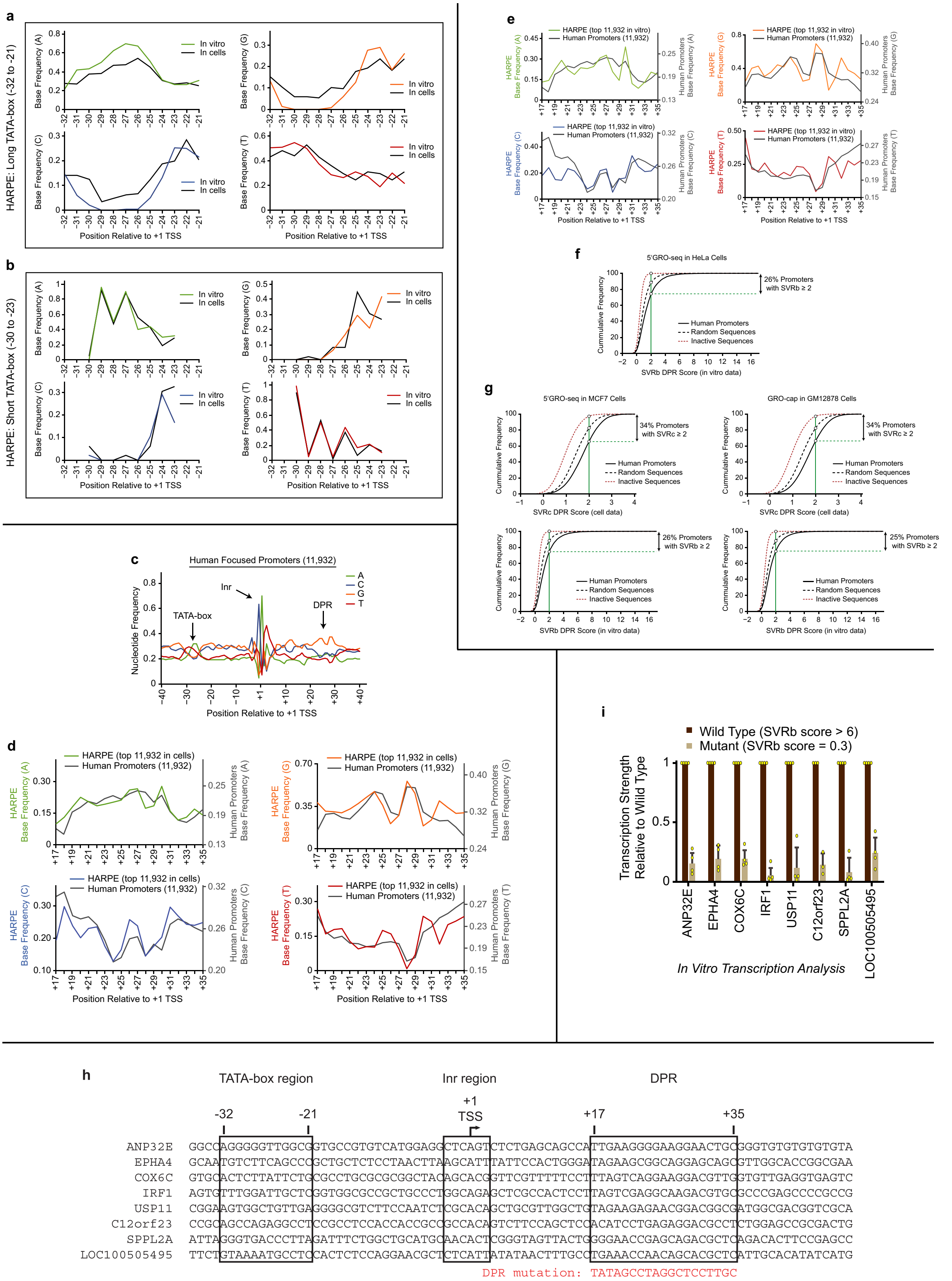
Transcription Strengths of HARPE Sequence Variants that Contain the Best 8-nt HOMER DPR Motif at Different Positions in the DPR Region

	+17	+18	+19	+20	+21	+22	+23	+24	+25	+26	+27	+28
Number of variants at each position	1787	2195	1948	1941	1898	2095	1792	1958	2013	2224	2027	2075
rho (SVRb score vs. observed):	0.86	0.84	0.86	0.82	0.86	0.91	0.83	0.80	0.94	0.95	0.89	0.79
rho (19-nt HOMER motif score vs. observed):	0.42	0.31	0.43	0.42	0.16	0.53	0.38	0.40	-0.03	0.71	0.31	0.25
rho (8-nt HOMER motif score vs. observed):	0.05	0.04	0.01	0.00	0.05	0.09	0.03	0.08	0.28	0.29	0.18	0.01

$$\text{Position Index at position } X = \frac{1}{200}\sum_{k=1}^{200} [\max - \min]_{\text{prediction score of A/C/G/T at position } X \text{ in sequence } k}$$

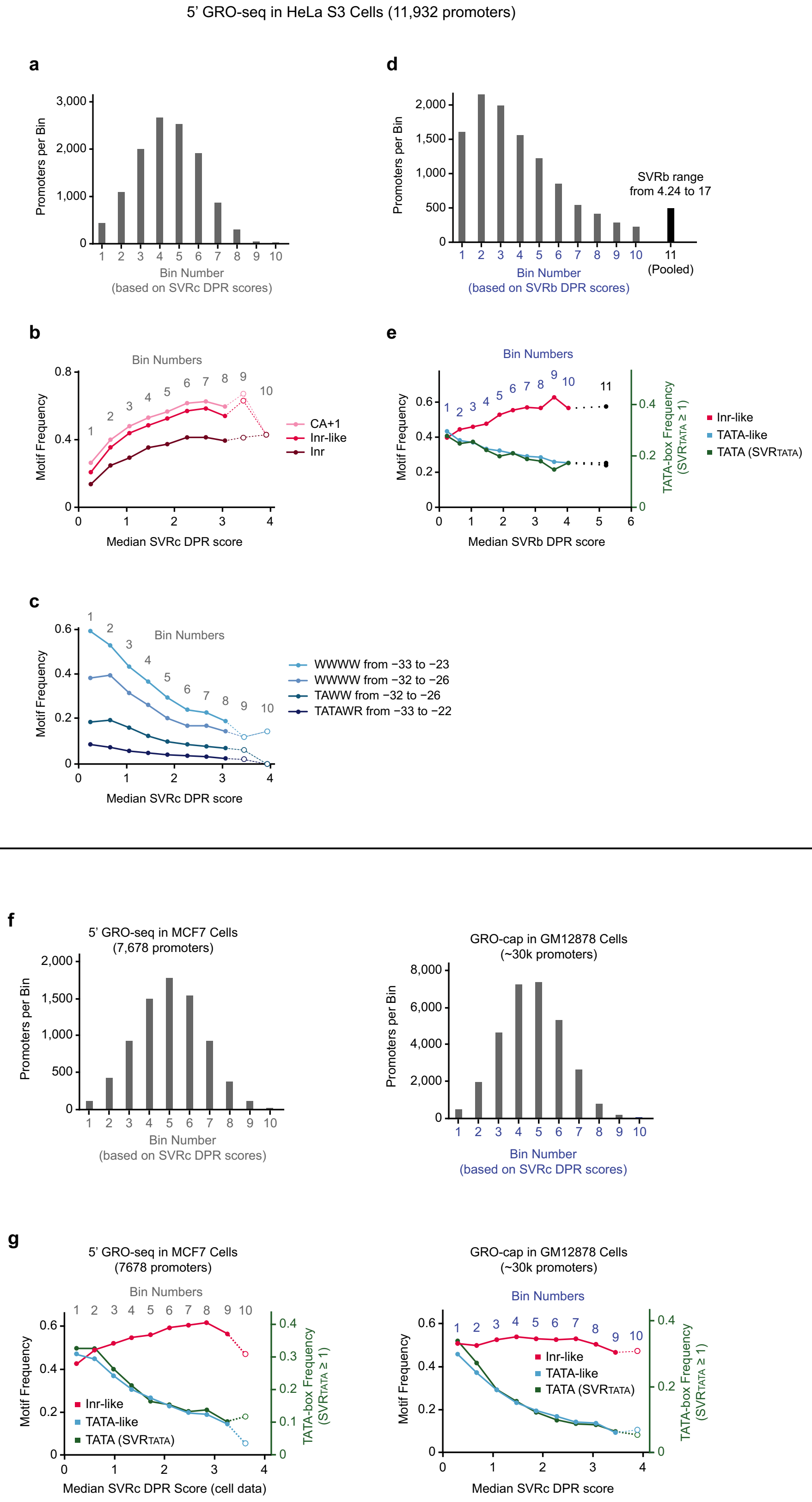
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Characterization of the properties of the SVR models and the generation of SVRtata for the TATA box and SVRc for the DPR with cell-based data. a–**c**, SVR models capture the preferred distances between the TSS and the DPR. **a**, The most significantly enriched 8-nt HOMER motif found in the top 0.1% of HARPE DPR variants (top) and its associated position-weight matrix (bottom). P-value associated with hypergeometric tests (one tailed). This 8-nt DPE-like motif closely resembles the *Drosophila* DPE consensus sequence[2,14]. Importantly, the DPE-like sequence is shorter than the DPR region and is therefore not at a fixed position. **b**, Positional preference analysis of the 8-nt motif in the top 0.1% HARPE DPR variants shows a preferred major position (74%) as well as a minor position (17%) that is 1 nt upstream of the major position. **c**, SVRb accurately predicts the transcription strength of sequence variants in all positions. This figure shows box-plot diagrams of the transcription strength for all variants within the HARPE dataset that contain the 8-nt motif at each position. The quality of the prediction at each position is indicated by Spearman's rank correlation coefficient (rho) between the observed transcription strength and SVRb score, HOMER motif score with the 19-nt DPR motif (shown in Extended Data Fig. 1e, f), or HOMER motif score with the 8-nt DPR motif (shown in **a**). The thick horizontal lines are the medians, and the lower and upper hinges are the first and third quartiles, respectively. Each upper (or lower) whisker extends from the upper (or lower) hinge to the largest (or lowest) value no further than 1.5 * IQR from the hinge. Data beyond the end of the whiskers (outlying points) are omitted from the box plot. All panels show a representative experiment ($n = 2$ biologically independent samples). **d**–**i**, Machine learning analysis of the HARPE TATA-box data yields an SVRTATA model for the TATA box. The HARPE data for the long TATA-box region (−32 to −21; Extended Data Figs. 1a, 2k–p, 8a, b) were subjected to SVR analysis. The resulting SVR models (derived from data generated in vitro or in cells) were termed SVRTATA. **d**, The SVRTATA model from HARPE data in cells is similar to that from HARPE data in vitro. The SVRTATA (in vitro) and SVRTATA (in cells) scores are compared by using 5000 independent test sequences that were not used in the training of the SVR. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. **e**, Comparison of SVRTATA scores and the observed transcription strengths of 5000 independent test sequences. These results are based on in vitro data. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. **f**, Comparison of HOMER motif scores and the observed transcription strengths of the same 5000 test sequences used in **e**. The position-weight matrices of the top 12-nt (left) or 8-nt (right) HOMER motifs (Extended Data Fig. 2n) were used to determine HOMER motif scores. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. **g**, Cumulative frequency of SVRTATA scores of natural human promoters in HeLa cells. Approximately 23% of 11,932 human promoters and 4% of 100,000 random sequences (61% average G/C content, as in human core promoters) have an SVRTATA (in vitro) score of at least 1 (marked with a green line), which corresponds to an active TATA box (Extended Data Fig. 5c). **h**, Cumulative frequency of SVRTATA scores of natural human promoters in MCF7 cells. Focused promoters identified in ref. [12] were used. Approximately 18% of 7,678 MCF7 promoters and 4% of 100,000 random sequences (61% average G/C content, as in human core promoters) have an SVRTATA (in vitro) score of at least 1 (marked with a green line), which corresponds to an active TATA box. **i**, Cumulative frequency of SVRTATA scores of natural human promoters in GM12878 cells. Focused promoters were identified as described in ref. [12] by using GRO-cap data in human GM12878 cells from ref. [37]. Approximately 15% of 30,643 GM12878 promoters and 4% of 100,000 random sequences (61% average G/C content, as in human core promoters) have an SVRTATA (in vitro) score of at least 1 (marked with a green line), which corresponds to an active TATA box. All panels show a representative experiment ($n = 2$ biologically independent samples). **j**, **k**, Most positions within the DPR have a moderate impact upon the overall SVR score. The influence of each position in the DPR on the model prediction score is shown by the value of the Position Index. The Position Index at position X is the average of the maximal magnitude of variation in (**j**) the SVR score or (**k**) the HOMER motif score with A, C, G or T at position X with 200 different DPR sequences that were randomly selected from natural human promoters. As a reference, the Web Logo for the top HOMER motif identified with the 0.1% most transcribed DPR sequences is also shown. **l**, **m**, SVRc model of the DPR with HARPE data generated in cells. **l**, HARPE libraries were transfected in cells, and normalized RNA tags were obtained. The SVRc (SVR from cell-based data) scores derived from these data correlate with measured transcription strengths in cells (with data that are independent of the SVRc training data) (representative experiment, $n = 2$ biologically independent samples). PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$. **m**, The SVRc models obtained from cells are reproducible. PCC, Pearson's correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$; rho, Spearman's rank correlation coefficient with two-tailed P-value $<2.2 \times 10^{-16}$.



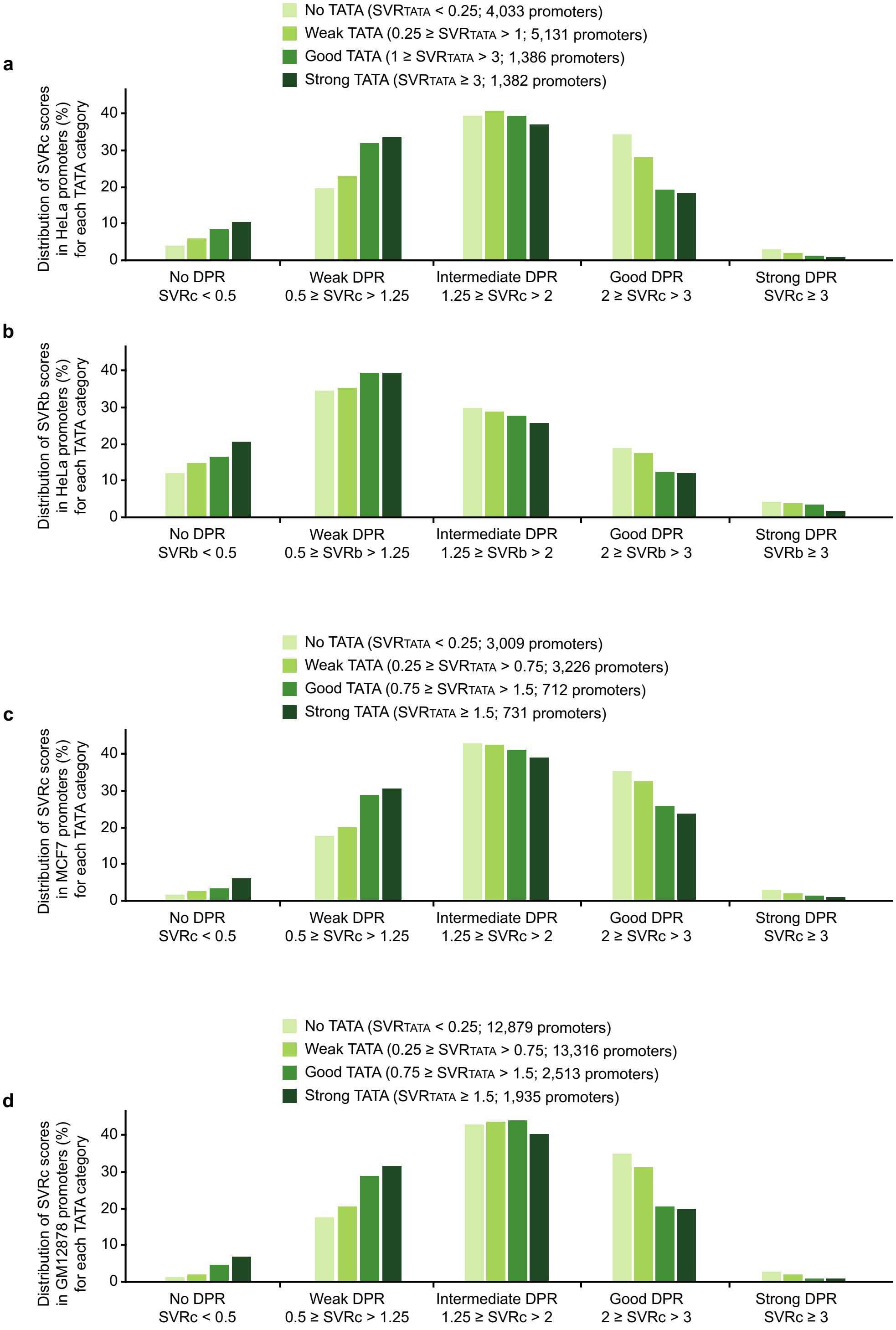
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Analysis of the HARPE TATA data as well as the DPR in natural human promoters. a, **b**, The nucleotide preferences of the top 0.1% most active TATA-box sequences in cells are similar to those of their in vitro counterparts. **a**, Long randomized TATA-box region (-32 to -21 relative to the +1 TSS). **b**, Short randomized TATA-box region (-30 to -23 relative to the +1 TSS). All panels show a representative experiment (*n* = 2 biologically independent samples). **c**, Distinct nucleotide preferences can be seen at the DPR in focused human promoters, which were identified as described in ref. [12] by using 5′GRO-seq data in HeLa cells[33]. **d**, The top ~2.5% (11,932) most active DPR sequences in cells, as assessed by HARPE, have nucleotide preferences that are similar to those seen in natural human core promoters in HeLa cells (representative experiment, *n* = 2 biologically independent samples). **e**–**g**, Relationship between natural human promoter sequences and HARPE data in vitro. **e**, The top ~2.5% (11,932) most active DPR sequences in vitro, as assessed by HARPE, have nucleotide preferences that are similar to those seen in natural human core promoters in HeLa cells. **f**, Cumulative frequency of SVRb DPR scores of natural human promoters. Approximately 26% of 11,932 human promoters (HeLa cells), 12% of 100,000 random sequences (61% average G/C content, as in human core promoters), and 0.4% of 10,000 inactive sequences (randomly selected from the 50% least active sequences in the HARPE assay; not used in the training of the SVR) have an SVRb score of at least 2 (marked with a green line), which corresponds to an active DPR (Extended Data Fig. 5a). **g**, Cumulative frequency of SVRc and SVRb DPR scores of natural human promoters in MCF7 and GM12878 cells. Approximately 34% of 7,678 MCF7 promoters, 34% of 30,643 GM12878 promoters, 17% of 100,000 random sequences (61% average G/C content, as in human core promoters), and 2.6% of 10,000 inactive sequences (randomly selected from the 50% least active sequences in the HARPE assay; not used in the training of the SVR) have an SVRc score of at least 2 (marked with a green line), which corresponds to an active DPR (Extended Data Fig. 5b). Approximately 26% of 7,678 MCF7 promoters, 25% of 30,643 GM12878 promoters, 12% of 100,000 random sequences (61% average G/C content, as in human core promoters), and 0.4% of 10,000 inactive sequences (randomly selected from the 50% least active sequences in the HARPE assay; not used in the training of the SVR) have an SVRb score of at least 2 (marked with a green line), which corresponds to an active DPR (Extended Data Fig. 5a). All panels show a representative experiment (*n* = 2 biologically independent samples). **h**, **i**, Analysis of the DPR in natural human promoters. **h**, Sequences of natural human promoters that contain DPR motifs with an SVRb score >6 and an SVRc score >2.5. The mutant DPR sequence has an SVRb score = 0.3 and an SVRc score = 0.3. **i**, Mutational analysis reveals DPR activity in different human promoters with SVRb DPR scores >6. In each of the mutant promoters, the wild-type DPR was substituted with a DNA sequence that has an SVRb DPR score of 0.3 (data are depicted as the mean with error bars denoting standard deviation, *n* = 3 or 4 biologically independent samples, as indicated by the points representing independent samples on the graph). The sequences of the tested promoters are shown in **f**. Promoter activity was measured by in vitro transcription followed by primer extension analysis of the TSSs. All P-values <0.01 (Student's *t*-test, two-tailed, paired). For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Analysis of the DPR and its relationship to the Inr and TATA box in active human promoters in different human cell lines. **a**–**e** Analysis of the DPR and its relationship to the Inr and TATA box in active human promoters in HeLa cells. **a**, Distribution of focused human promoters derived from HeLa cells in increasing SVRc DPR score bins. Bins 9 and 10 have less than 100 promoters. **b**, The frequencies of occurrence of the Inr and Inr-like sequences in different bins of promoters with increasing SVRc DPR scores. The Inr-like sequence is as defined previously[12]. **c**, The frequencies of occurrence of the TATA box and TATA-like sequences decrease as the SVRc DPR score increases. **d**, Distribution of focused human promoters in increasing SVRb DPR score bins. Promoters with SVRb scores between 4.24 and 17 were combined together in bin 11. **e**, The frequencies of occurrence of Inr-like sequences, TATA-like sequences, and TATA-box motifs (as assessed with SVRTATA ≥ 1; Extended Data Fig. 5c) in different bins of promoters with increasing SVRb DPR scores. The Inr-like and TATA-like sequences are as defined previously[12]. In **b** and **c**, bins with less than 100 promoters are indicated with open circles and are connected by dashed lines. In **e**, bin 11 is shown in black circles connected by dashed black lines. All panels show a representative experiment (*n* = 2 biologically independent samples). **f**, **g**, Analysis of the DPR and its relationship to the Inr and TATA box in active human promoters in MCF7 and GM12878 cells. **f**, Distribution of focused human promoters in increasing SVRc DPR score bins. For each cell line, bin 10 has less than 100 promoters. MCF7 focused promoters are described in ref. [12]. GM12878 focused promoters were identified as described in ref. [12] by using GRO-cap data in human GM12878 cells from ref. [37]. **g**, The frequencies of occurrence of Inr-like sequences, TATA-like sequences, and TATA-box motifs (as assessed with SVRTATA ≥ 1; Extended Data Fig. 5c) in different bins of promoters with increasing SVRc DPR scores. The Inr-like and TATA-like sequences are as defined previously[12]. Bins with less than 100 promoters are indicated with open circles and are connected by dashed lines. All panels show a representative experiment (*n* = 2 biologically independent samples).



Extended Data Fig. 10 | Distribution of SVR DPR scores for human promoters in relation to their SVRtata scores. Human promoters were divided into four groups according to their SVRTATA score. For each TATA box category, the distribution of SVR DPR scores is shown for each of five classes of promoters (no DPR, weak DPR, intermediate DPR, good DPR, and strong DPR). **a**, Human focused promoters obtained from HeLa cells[12,33] analysed with SVRTATA and SVRc. **b**, Human focused promoters obtained from HeLa cells analysed with SVRTATA and SVRb. **c**, Human focused promoters obtained from MCF7 cells[12] analysed with SVRTATA and SVRc. **d**, Human focused promoters obtained from GM12878 cells[37] analysed with SVRTATA and SVRc. Focused promoters were identified as described in ref. [12] by using GRO-cap data in human GM12878 cells from ref. [37]. All panels show a representative experiment ($n = 2$ biologically independent samples).

nature research

Corresponding author(s): Kadonaga, James T.

Last updated by author(s): Jun 7, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data from radiolabeled samples were collected with a GE Amersham Typhoon 5 and the Amersham™ Typhoon™ control software v1.1. Illumina sequencing was conducted on a HiSeq 4000 or a Novseq 6000.
Data analysis	Quantification of radiolabeled samples was measured with Fiji v1.52i. All other analyses were performed on R v3.6.1 through Rstudio v1.1.463 with packages ggplot2 v3.2.1, tidyr v1.0.0, e1071 v1.7-2 and rlist v0.4.6.1. Adobe Illustrator CS v11.0.0 for building figures.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data supporting the findings of this study are available within the paper and its supplementary information files. The HARPE data are publicly available at the Gene Expression Omnibus (GEO; accession number, GSE139635). The 5’-GRO-seq files (GSE63872) and (GSE90035) and the GRO-cap files (GSM1480321) were obtained from the Gene Expression Omnibus website.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculation was performed. Sample sizes were chosen in a manner commensurate with similar previous studies.
Data exclusions	Accurate NGS reads were selected based on the criteria described in the methods section. No other data were excluded from the analysis. For experiments testing individual variants/promoters, one out of 31 promoter pairs (WT vs mutant) tested was an outlier (Dixon's Q-test, two-tailed, 96% confidence; criterion not pre-established) and removed from the analysis. Including this data point in the analysis would not change the conclusions of the experiment.
Replication	For experiments testing individual variants/promoters, three or more biological replicates were performed. All attempts at replication were successful. For HARPE experiments and experiments involving Sarkosyl, two biological replicates were performed. All HARPE experiments were successful.
Randomization	Randomization was not relevant for this study as it did not involve allocation into experimental groups.
Blinding	Blinding was not relevant for this study as it did not involve allocation into experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HeLa cells were a gift from Anjana Rao (La Jolla Institute for Immunology), originally obtained from the ATCC.
Authentication	Cell lines were not authenticated.
Mycoplasma contamination	HeLa cells were negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Structures of α-synuclein filaments from multiple system atrophy

https://doi.org/10.1038/s41586-020-2317-6

Received: 5 February 2020

Accepted: 22 April 2020

Manuel Schweighauser[1,9], Yang Shi[1,9], Airi Tarutani[2,3], Fuyuki Kametani[2], Alexey G. Murzin[1], Bernardino Ghetti[4], Tomoyasu Matsubara[5], Taisuke Tomita[3], Takashi Ando[6], Kazuko Hasegawa[7], Shigeo Murayama[5], Mari Yoshida[8], Masato Hasegawa[2], Sjors H. W. Scheres[1,10]✉ & Michel Goedert[1,10]✉

Synucleinopathies, which include multiple system atrophy (MSA), Parkinson's disease, Parkinson's disease with dementia and dementia with Lewy bodies (DLB), are human neurodegenerative diseases[1]. Existing treatments are at best symptomatic. These diseases are characterized by the presence of, and believed to be caused by the formation of, filamentous inclusions of α-synuclein in brain cells[2,3]. However, the structures of α-synuclein filaments from the human brain are unknown. Here, using cryo-electron microscopy, we show that α-synuclein inclusions from the brains of individuals with MSA are made of two types of filament, each of which consists of two different protofilaments. In each type of filament, non-proteinaceous molecules are present at the interface of the two protofilaments. Using two-dimensional class averaging, we show that α-synuclein filaments from the brains of individuals with MSA differ from those of individuals with DLB, which suggests that distinct conformers or strains characterize specific synucleinopathies. As is the case with tau assemblies[4–9], the structures of α-synuclein filaments extracted from the brains of individuals with MSA differ from those formed in vitro using recombinant proteins, which has implications for understanding the mechanisms of aggregate propagation and neurodegeneration in the human brain. These findings have diagnostic and potential therapeutic relevance, especially because of the unmet clinical need to be able to image filamentous α-synuclein inclusions in the human brain.

A causal link between α-synuclein assembly and disease was established by the findings that missense mutations in *SNCA* (the gene that encodes α-synuclein) and multiplications of this gene give rise to rare inherited forms of Parkinson's disease and Parkinson's disease with dementia[3,10]. Some mutations also cause DLB. The missense mutations in *SNCA* that result in G51D[11,12] and A53E[13] substitutions can give rise to atypical synucleinopathies, with a mixture of Parkinson's disease and MSA pathologies. Sequence variation in the regulatory region of *SNCA* is associated with an increased expression of α-synuclein and a heightened risk of developing idiopathic Parkinson's disease[14], which accounts for over 90% of cases of this disease.

MSA is a sporadic synucleinopathy of adult onset, with symptoms of parkinsonism, cerebellar ataxia and autonomic failure[15–17]. Cases of MSA are classified as MSA-P, which show predominant parkinsonism caused by striatonigral degeneration, and MSA-C, which show cerebellar ataxia associated with olivopontocerebellar atrophy. Autonomic dysfunction is common to both subtypes. In neuropathological terms, MSA is defined by regional nerve cell loss and the presence of abundant filamentous α-synuclein inclusions in oligodendrocytes: glial cytoplasmic inclusions, known as Papp–Lantos bodies[18–21]. Smaller numbers of α-synuclein inclusions are also present in nerve cells[22]. The mean duration of the disease is 6–10 years, but survival times of 18–20 years have been reported. The late appearance of autonomic dysfunction correlates with prolonged survival[23].

α-Synuclein is a 140-amino-acid protein, over half of which (residues 7–87) consists of 7 imperfect repeats, with the consensus sequence KTKEGV. These residues encompass the lipid-binding domain[24]. The repeats partially overlap with a hydrophobic region (residues 61–95) known as the non-β-amyloid component[25], which is necessary for the assembly of recombinant α-synuclein into filaments[26]. The carboxy-terminal region (residues 96–140) is negatively charged, and its truncation results in increased filament assembly[27]. Upon assembly, recombinantly expressed α-synuclein undergoes conformational changes and takes on a cross-β structure that is characteristic of amyloid[28,29]. The cores of α-synuclein filaments extracted from the cerebellum of patients with MSA, or assembled from recombinant protein in vitro, encompass around 70 amino acids that extend approximately from residue 30 to residue 100[30].

Seeded assembly of α-synuclein, propagation of inclusions and nerve cell death have been demonstrated in a variety of systems[31–34]. Assemblies of recombinant α-synuclein with different morphologies

---

[1]MRC Laboratory of Molecular Biology, Cambridge, UK. [2]Department of Brain and Neurosciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan. [3]Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan. [4]Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA. [5]Department of Neuropathology, Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan. [6]Department of Neurology, Nagoya University Graduate School of Medicine, Nagoya, Japan. [7]Division of Neurology, Sagamihara National Hospital, Sagamihara, Japan. [8]Institute for Medical Science of Aging, Aichi Medical University, Nagakute, Japan. [9]These authors contributed equally: Manuel Schweighauser, Yang Shi. [10]These authors jointly supervised this work: Sjors H. W. Scheres, Michel Goedert. ✉e-mail: scheres@mrc-lmb.cam.ac.uk; mg@mrc-lmb.cam.ac.uk

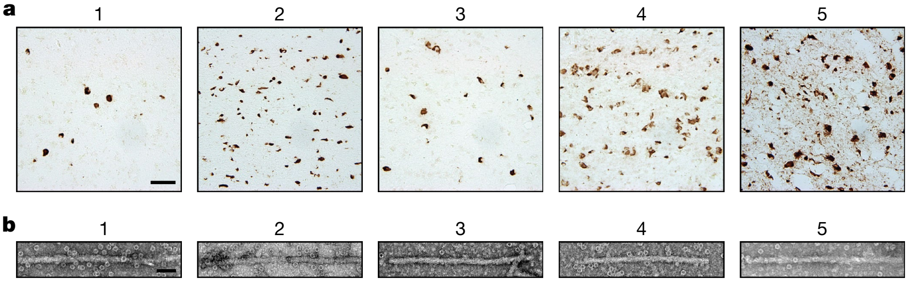


Fig. 1 | Filamentous α-synuclein pathology of MSA. a, Staining of neuronal and glial inclusions in the putamen in MSA cases 1–5 by the pS129 antibody (brown). Scale bar, 50 μm. **b**, Negative-stain electron micrographs of filaments from the putamen in MSA cases 1–5. Spherical densities probably correspond to ferritin that purified with the filaments. Scale bar, 50 nm.

have displayed distinct seeding capacities[35]. Moreover, α-synuclein from glial cytoplasmic inclusions has previously been reported to be approximately three orders of magnitude more potent than α-synuclein from Lewy bodies in seeding the aggregation of α-synuclein[36]. Indirect evidence has also suggested that distinct conformers of assembled α-synuclein may characterize MSA and disorders with Lewy pathology[37–43]. Solubility in sodium dodecyl sulfate (SDS) distinguishes α-synuclein filaments of MSA from those of DLB[44].

Neuropathological characteristics

We used sarkosyl to extract filaments from the putamen of five individuals with a neuropathologically confirmed diagnosis of MSA (hereafter referred to as MSA cases 1–5). In MSA cases 1, 2, 3 and 5, filaments were also extracted from the frontal cortex; the same was true of the cerebellum for MSA case 1. Most sarkosyl-insoluble α-synuclein phosphorylated at S129 was soluble in SDS. More than 90% of α-synuclein inclusions are phosphorylated at S129[45]. For MSA case 1, the individual was diagnosed as MSA-P and had an age at death of 85 years; for MSA cases 2, 3, 4 and 5, the individuals were diagnosed as MSA-C and had ages at death of 68, 59, 64 and 70 years, respectively. The disease durations were 9, 18, 9, 10 and 19 years for MSA cases 1, 2, 3, 4 and 5, respectively.

The abundant glial cytoplasmic inclusions and neuronal inclusions were stained by an antibody specific for α-synuclein phosphorylated at S129 (hereafter referred to as antibody pS129) (Fig. 1a, Extended Data Fig. 1a). By negative-stain electron microscopy, all five cases of MSA showed a majority of twisted filaments, which had a diameter of 10 nm and a periodicity of 80–100 nm (Fig. 1b, Extended Data Fig. 1b). Immunogold negative-stain electron microscopy with the anti-α-synuclein antibody PER4 showed decoration of MSA filaments (Extended Data Fig. 1c, d), consistent with previous findings[20]. Immunoblotting sarkosyl-insoluble material from the putamen with the antibodies Syn303 and PER4 revealed evidence of monomeric α-synuclein and high-molecular-weight aggregates (Extended Data Fig. 1e). Truncated α-synuclein was also present. When the antibody pS129 was used, full-length α-synuclein was the predominant species. Consistent with the results of immunostaining (Fig. 1a), in MSA cases 1 and 3 the putamen contained lower levels of α-synuclein than it did in MSA cases 2, 4 and 5.

We observed the seeded aggregation of expressed wild-type human α-synuclein in SH-SY5Y cells after the addition of sarkosyl-insoluble seeds from the putamen of MSA cases 1–5 (Extended Data Fig. 2). Seeds from MSA case 3 were the most potent, and those from MSA case 2 were least effective at inducing seeded aggregation. Seeds from MSA cases 1, 4 and 5 had intermediate seeding potencies.

Two types of MSA filament

We imaged the sarkosyl-insoluble filaments using cryo-electron microscopy (cryo-EM) (Extended Data Fig. 3). These filaments looked identical upon visual inspection of the micrographs, but reference-free 2D class averaging revealed two types of filament (Extended Data Fig. 3b, d). Type I filaments were less symmetrical than the type II filaments. In the putamen, the ratios of type I to type II filaments were 80:20 in MSA case 1 and 20:80 in MSA case 2. MSA cases 3 and 4 had mostly type I filaments, and MSA case 5 had only type II filaments (Fig. 2a, b).

This suggests that the duration of MSA may correlate with the ratio of filament types in putamen, but additional cases of disease are required to establish this more firmly. What is true for the putamen may not be true of α-synuclein filaments from other affected regions of the brain. We identified predominantly type I filaments in the putamen in MSA cases 1 and 3 (Fig. 2), whereas we found almost exclusively type II filaments in the cerebellum in MSA case 1 and in the frontal cortex in MSA cases 2, 3 and 5 (Extended Data Fig. 3d). It remains to be seen whether the MSA type I and type II filaments are common to both nerve and glial cells.

Protofilaments adopt extended folds

We determined the cryo-EM structures of MSA type I and type II filaments from the putamen to resolutions sufficient for de novo atomic modelling (Fig. 2, Extended Data Table 1). The best structures were resolved to a resolution of 2.6 Å for type I filaments in MSA case 1, and a resolution of 3.1 Å for type II filaments in MSA case 2 (Extended Data Fig. 4). Type I and type II filaments are each made of two protofilaments, which consist of an extended N-terminal arm and a compact C-terminal body (Fig. 2, Extended Data Fig. 5). Both the type I and the type II filaments are asymmetrical. The larger protofilament of the type I filaments (PF-IA) comprises residues G14–F94 of α-synuclein, and the smaller protofilament (PF-IB) consists of residues K21–Q99 (Fig. 2c). For type II filaments, PF-IIA and PF-IIB comprise residues G14–F94 and G36–Q99, respectively (Fig. 2d). Protofilament folds differ from each other within and between filament types. MSA type I and type II filaments are thus collectively made of four distinct protofilaments (Figs. 2, 3a).

PF-IA comprises 12 β-strands. The N-terminal arm of PF-IA consists of a cross-β hairpin (residues G14–G31) and an extended one-layered L-shaped motif at residues K32–K45. The C-terminal body of PF-IA adopts a three-layered L-shaped motif. The outer layer (residues E46–V66) is the longest of these layers, and packs against the outside of the central layer (residues G67–E83). A salt bridge between E46 and K80 stabilizes this interaction. The shorter inner layer (residues G84–F94) packs against the inside of the central layer. Glycine-rich turns connect the layers. PF-IB comprises 10 β-strands. The N-terminal arm of PF-IB consists only of a cross-β hairpin at residues G25–K45. The three-layered L-shaped motif of the C-terminal body of PF-IB is topologically similar to that in PF-IA. Nevertheless, the two motifs differ in structure—most notably in the packing of the inner layer against the central layer by the residues that follow G86. The body of PF-IA ends at F94, whereas the body of PF-IB extends to Q99.

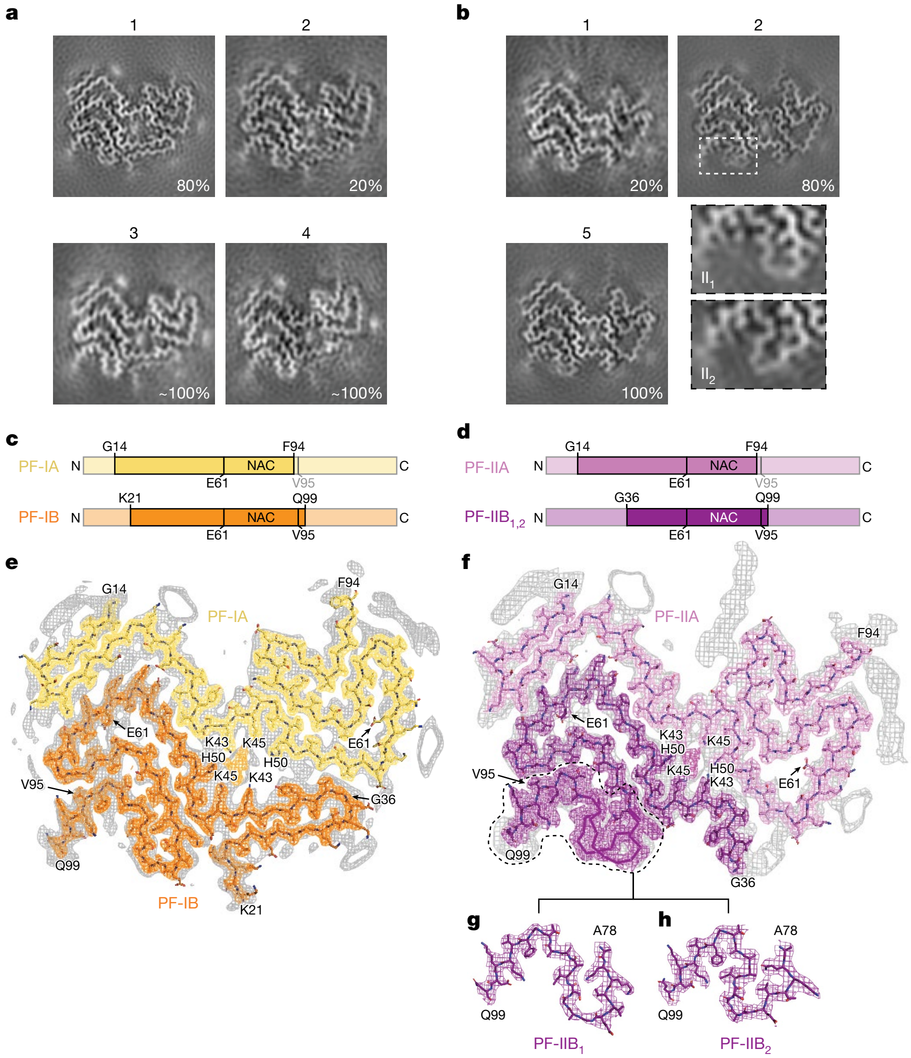


Fig. 2 | Cryo-EM maps and atomic models of type I and type II filaments of α-synuclein from MSA. a, b, Cryo-EM maps of type I filaments from the putamen in MSA cases 1, 2, 3 and 4 (**a**) and of type II filaments from the putamen in MSA cases 1, 2 and 5 (**b**). For MSA case 2, enlarged views of the different regions in type II_1 and type II_2 filaments are also shown. **c, d**, Schematic of the primary structure of human α-synuclein, indicating the cores of PF-IA, PF-IB, PF-IIA and PF-IIB. The non-β-amyloid component (NAC) domain (residues 61–95) is also shown. C, C terminus; N, N terminus. **e, f**, Sharpened, high-resolution cryo-EM maps of type I (**e**) and type II (**f**) filaments of α-synuclein from MSA, with overlaid atomic models. Unsharpened, 4.5 Å low-pass-filtered maps are in grey. The high-resolution maps show weaker densities that extend from the N- and C-terminal regions, a peptide-like density in PF-IIA, and weaker densities that border the solvent-exposed chains of K32 and K34 in PF-IA, PF-IB and PF-IIA. Weaker densities that border the solvent-exposed chains of K58 and K60 in PF-IA and PF-IIA are also present. **g, h**, Cryo-EM structures of A78–Q99 of PF-IIB, illustrating heterogeneity ($PF\text{-}IIB_1$ and $PF\text{-}IIB_2$). There is strong density at the protofilament interfaces of type I and type II filaments, which is surrounded by the side chains of K43, K45 and H50 from each protofilament.

PF-IIA also comprises 12 β-strands and spans residues G14–F94. The PF-IA and PF-IIA protofilaments have similar N-terminal arms (Fig. 3). Although the C-terminal body of PF-IIA adopts a three-layered L-shaped motif, its conformation differs from that of PF-IA. In type I protofilaments residues G47–V52 from the outer layer pack against residues A76–K80 from the central layer, whereas in PF-IIA this packing is shifted by two residues and involves residues V74–A78 in the central layer. This creates a sizeable cavity between the central layer and the L-shaped bend at E57 in the outer layer. This shift also increases the distance between the Cα atoms of E46 and K80 by 5 Å, but a salt bridge may still form between their side chains. PF-IIB is the smallest protofilament core and comprises 9 β-strands. The N-terminal arm of PF-IIB is made of a single L-shaped conformation at residues G36–K45. The C-terminal body of PF-IIB forms a three-layered L-shaped motif, which exists in two conformations: $PF\text{-}IIB_1$, which is virtually identical to PF-IB, and $PF\text{-}IIB_2$, which has a different backbone conformation at residues T81–A90 (Fig. 2f–h). On the basis of the number of classified helical segments, the ratio of type II filaments containing $PF\text{-}IIB_1$ (type II_1) to type II filaments containing $PF\text{-}IIB_2$ (type II_2) was 20:80 (Fig. 2b, Extended Data Table 1).

Filaments enclose additional molecules

In MSA type I and type II filaments, two non-identical protofilaments pack against each other through an extended interface that forms a large cavity surrounded by the side chains of K43, K45 and H50 from each protofilament (Fig. 2, Extended Data Figs. 5, 6a, b). This cavity encloses an additional strong density that is not connected to the protein density (Fig. 2, Extended Data Fig. 4). The chemical nature

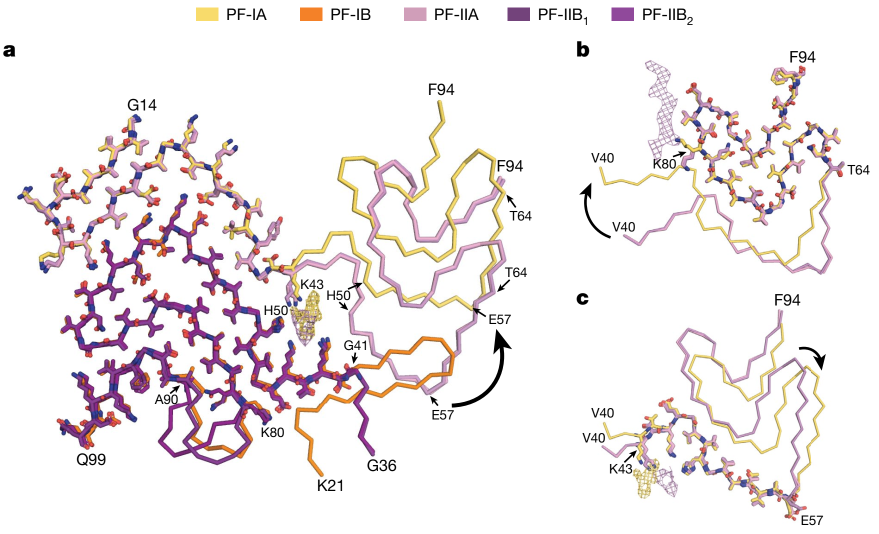


Fig. 3 | Comparison of the protofilament folds in α-synuclein from MSA. **a**, Overlay of the structures of PF-IA, PF-IB, PF-IIA and PF-IIB from MSA. The black arrow indicates the direction of the conformational change that occurs at K43 of PF-IA and PF-IIA. **b**, **c**, Three-layered L-shaped motifs of PF-IA (yellow) and PF-IIA (pink) are aligned, on the basis of structural similarities between T64–F94 (**b**) and T44–E57 (**c**). Black arrows indicate the direction of the conformational change that occurs at T64 (**b**) or E57 (**c**) of PF-IA and PF-IIA. The peptide-like density in PF-IIB is shown as a pink mesh.

of this density remains to be established. The observations that it is disconnected from the density of the α-synuclein polypeptide chains and that it would need to compensate four positive charges for every β-sheet rung suggest that this density is non-proteinaceous. The cavity is larger in type I filaments than in type II filaments and contains additional, smaller densities between H50, G51 and A53 of PF-IA, and V37 and Y39 of PF-IB. Although we used sarkosyl to extract filaments, the central cavity is not large enough to accommodate one sarkosyl molecule per rung. Moreover, the negative charge of the headgroup of sarkosyl (−1) cannot compensate for the positive charge (+4 per rung) of the central cavity, and the polar nature of the cavity is not compatible with the fatty-acid tail of sarkosyl.

Besides the density in the large cavity at the interface of the protofilaments, several other densities are visible at lower intensities. At the N and C termini of the ordered cores of all four protofilaments, fuzzy densities probably correspond to less-well-ordered extensions of the core. The longest extensions are seen for PF-IA and PF-IIA. Unlike PF-IA, a peptide-like density of unknown identity is packed against residues K80–E83 of PF-IIA. This density may correspond to an extension of the C- terminus of α-synuclein in PF-IIA, or to an unknown protein that is bound to the filament core. Additional unconnected densities are observed in front of pairs of lysines on the exterior of the filaments—that is, in front of K32 and K34 of PF-IA, PF-IB and PF-IIA, as well as in front of K58 and K60 of PF-IA and PF-IIA. Similar densities have previously been observed in front of pairs of lysines on the exterior of tau filaments from Alzheimer's disease[4,6], Pick's disease[5], chronic traumatic encephalopathy[7] and corticobasal degeneration[9], although the molecules that form these densities remain unknown.

In the structures of MSA type I and type II filaments, residues G51 and A53 of α-synuclein form part of the interfaces of the protofilaments, and are located close to K43, K45 and H50. The mutations in *SNCA* that lead to G51D and A53E substitutions are the only known disease-causing mutations that increase the negative charge of α-synuclein[11–13]. All four protofilaments of the MSA filaments can accommodate the side chains of D51 or E53 without substantial structural changes (Extended Data Fig. 6c, d). The presence of D51 and A53 may thus give rise to similar type I and type II filament structures. However, the changes in charge of the residues that surround the central cavity may lead to a different molecular composition of the additional density in cases of MSA with G51D and A53E substitutions, as compared to sporadic MSA.

Using mass spectrometry of sarkosyl-insoluble α-synuclein from the putamen, we found that N-terminal acetylation, C-terminal truncation and ubiquitination of K6 and K12 were common to MSA cases 1–5. In the sequences of the filament cores, K21 was ubiquitinated in all cases. Despite having identical structures of type I and type II filaments, in only some cases of MSA did sarkosyl-insoluble α-synuclein also show ubiquitination of K23, K60 and K80, acetylation of K21, K23, K32, K34, K45, K58, K60, K80 and K96, as well as phosphorylation of Y39, T59, T64, T72 and T81. With the exception of ubiquitination of K80, the percentage of α-synuclein molecules modified at a given residue was low, which suggests that these post-translational modifications are not responsible for the additional densities in the cryo-EM maps. Some of these modifications have previously been described[1], but others are newly described here.

Ubiquitination of K80 was detected in sarkosyl-insoluble α-synuclein in MSA cases 2 and 5, which show a preponderance of type II filaments. This bulky post-translational modification, which is compatible with the structure of PF-IIA, clashes with the surroundings of the K80 side chain in PF-IA, PF-IB and PF-IIB. Moreover, one end of the peptide-like density—which is specific to type II filaments—is located next to K80 of PF-IIA. This density, which does not appear to be connected to the side chain of K80, may consist of a mixture of different sequences and ubiquitination might possibly have a role. Phosphorylation of T72 may favour PF-IIA over PF-IA. The side chain of T72 is buried in PF-IA, whereas it borders a large cavity between the outer and central layers in PF-IIA. Phosphorylation of T81 may distinguish PF-IIB$_1$ and PF-IIB$_2$, as the side chain of this residue is buried in PF-IIB$_1$ but is solvent-exposed in PF-IIB$_2$. Post-translational modifications in only one protofilament may favour the formation of asymmetrical type I and type II filaments. Thus, in the structures of PF-IA and PF-IIA, the side chain of K60 is solvent-exposed and can carry a bulky modification. By contrast, in the structures of PF-IB and PF-IIB this side chain is buried in the interfaces between protofilaments.

DLB filaments

Our results show that α-synuclein filaments adopt the same structures in different individuals with MSA. Similar observations have previously been made for tau filaments from the brains of individuals with Alzheimer's disease[4,6], Pick's disease[5], chronic traumatic encephalopathy[7] and corticobasal degeneration[9]. Tau filaments adopt an identical fold in individuals with the same disease, but different tauopathies are characterized by distinct folds. To assess whether the same is true of synucleinopathies, we used cryo-EM to examine α-synuclein filaments that were isolated from the brains of three individuals with a neuropathologically confirmed diagnosis of DLB.

In the frontal cortex and amygdala, abundant Lewy bodies and Lewy neurites were stained by the pS129 antibody (Extended Data Fig. 7a). Following sarkosyl extraction, α-synuclein filaments from the brains of individuals with DLB did not appear to twist and were thinner than those from the brains of individuals with MSA (Extended Data Fig. 7b, d). Similar differences between α-synuclein filaments from the brains of individuals with Lewy pathology and those with MSA have previously been described[20,38]. Unlike MSA, most sarkosyl-insoluble α-synuclein phosphorylated at S129 from the brains of individuals with DLB was SDS-insoluble, consistent with previous findings[44]. Immunogold negative-stain electron microscopy with the antibody PER4 showed decoration of DLB filaments (Extended Data Fig. 7c), consistent with previous findings[46]. The lack of twist precluded the determination of the three-dimensional structure of α-synuclein filaments from DLB by cryo-EM, but on the basis of reference-free 2D class averages (Extended Data Fig. 7e) we conclude that the structures of α-synuclein filaments of DLB are different from those of MSA.

Synthetic filaments

We next compared the structures of filaments from the brains of individuals with MSA with those assembled in vitro from recombinant wild-type and mutant α-synucleins[47–54] (Extended Data Fig. 8). The largest differences are in the extended sizes of the MSA protofilaments, and in the asymmetrical packing of these protofilaments. None of the recombinant α-synuclein filaments contain the long N-terminal arms of MSA filaments and most recombinant filaments are made of either one protofilament or two identical protofilaments related by helical symmetry.

As with all of the MSA protofilaments, some recombinant α-synuclein protofilaments also contain three-layered L-shaped motifs (Extended Data Fig. 9). One feature that the recombinant α-synuclein filaments with the three-layered L-shaped motif have in common is that they were assembled in the presence of polyanions, such as phosphate, or of chaotropic negatively charged ions, such as bromine[47–50]. It has previously been suggested that the additional densities in front of the side chains of K43 and K45 from one protofilament, and of K58 from the other, correspond to phosphate ions[48,51]. This raises the possibility that the additional density in the cavity at the protofilament interface of MSA filaments may also consist of molecules that contain phosphate groups. Unlike in the recombinant α-synuclein structures, the pseudo-symmetric cavity in MSA type II filaments can accommodate approximately two phosphate groups per β-sheet rung, consistent with the size of the density. In type I filaments, the central cavity is more open on one side and can therefore accommodate a larger density.

The finding that the structures of α-synuclein filaments from MSA differ from those of assembled recombinant proteins is consistent with a previous observation that inhibitors of α-synuclein assembly affect aggregation by MSA and recombinant filament seeds in different ways[55]. It is also reminiscent of similar findings for tau filaments[4–9], even though marked differences exist between tau and α-synuclein. Recombinant tau requires cofactors to form filaments in vitro, whereas the assembly of recombinant α-synuclein proceeds in the absence of cofactors. Moreover, α-synuclein exists as a single protein of 140 amino acids, whereas 6 tau isoforms—which ranging in size from 352 to 441 amino acids—are expressed in the adult human brain; the isoform composition of filaments varies among some tauopathies[56]. However, as was the case with recombinant α-synuclein filaments[47–54], the structures of heparin-induced filaments of recombinant tau[8] differed from those present in disease[4–7,9]. In both cases, in vitro-assembled filaments were smaller and adopted topologically simpler conformations.

Outlook

Here we establish the presence of two types of α-synuclein filament in MSA, and suggest that different conformers or strains of assembled α-synuclein exist in DLB. To understand the causes and spreading of α-synuclein pathology as well as the distinguishing characteristics of synucleinopathies, it will be important to identify the mechanisms of seed formation and subsequent assembly. The presence of post-translational modifications in assembled α-synuclein is well-established, but their relevance for assembly is not understood[1]. In addition, the structures of α-synuclein filaments in MSA reveal the presence of non-proteinaceous molecules, reminiscent of findings in tauopathies[7,9]. It will be important to identify the chemical nature of these molecules and to study their effects—alone or in conjunction with post-translational modifications—on α-synuclein and tau assembly. Understanding the structural specificity of filament assembly in disease will facilitate the development of tracers for imaging filamentous amyloid assemblies of α-synuclein in the brain, and of molecules that prevent, inhibit and reverse filament formation.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2317-6.

1. Goedert, M., Jakes, R. & Spillantini, M. G. The synucleinopathies: twenty years on. *J. Parkinsons Dis*. **7**, S51–S69 (2017).
2. Spillantini, M. G. et al. α-Synuclein in Lewy bodies. *Nature* **388**, 839–840 (1997).
3. Polymeropoulos, M. H. et al. Mutation in the α-synuclein gene identified in families with Parkinson's disease. *Science* **276**, 2045–2047 (1997).
4. Fitzpatrick, A. W. P. et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
5. Falcon, B. et al. Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* **561**, 137–140 (2018).
6. Falcon, B. et al. Tau filaments from multiple cases of sporadic and inherited Alzheimer's disease adopt a common fold. *Acta Neuropathol*. **136**, 699–708 (2018).
7. Falcon, B. et al. Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568**, 420–423 (2019).
8. Zhang, W. et al. Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer's and Pick's diseases. *eLife* **8**, e43584 (2019).
9. Zhang, W. et al. Novel tau filament fold in corticobasal degeneration. *Nature* **580**, 283–287 (2020).
10. Singleton, A. B. et al. α-Synuclein locus triplication causes Parkinson's disease. *Science* **302**, 841 (2003).
11. Kiely, A. P. et al. α-Synucleinopathy associated with G51D *SNCA* mutation: a link between Parkinson's disease and multiple system atrophy? *Acta Neuropathol*. **125**, 753–769 (2013).
12. Kiely, A. P. et al. Distinct clinical and neuropathological features of G51D *SNCA* mutation cases compared with *SNCA* duplication and H50Q mutation. *Mol. Neurodegener*. **10**, 41 (2015).
13. Pasanen, P. et al. Novel α-synuclein mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathology. *Neurobiol. Aging* **35**, 2180.e1–2180. e5 (2014).
14. Nalls, M. A. et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet*. **46**, 989–993 (2014).
15. Dejerine, J. & Thomas, A. L'atrophie olivo-ponto-cérébelleuse. *Nouv. Iconogr. Salpêtrière* **13**, 330–370 (1900).
16. Graham, J. G. & Oppenheimer, D. R. Orthostatic hypotension and nicotine sensitivity in a case of multiple system atrophy. *J. Neurol. Neurosurg. Psychiatry* **32**, 28–34 (1969).
17. Quinn, N. Multiple system atrophy – the nature of the beast. *J. Neurol. Neurosurg. Psychiatry* **52** (suppl.), 78–89 (1989).
18. Papp, M. I., Kahn, J. E. & Lantos, P. L. Glial cytoplasmic inclusions in the CNS of patients with multiple system atrophy (striatonigral degeneration, olivopontocerebellar atrophy and Shy–Drager syndrome). *J. Neurol. Sci*. **94**, 79–100 (1989).
19. Wakabayashi, K., Yoshimoto, M., Tsuji, S. & Takahashi, H. α-synuclein immunoreactivity in glial cytoplasmic inclusions in multiple system atrophy. *Neurosci. Lett*. **249**, 180–182 (1998).
20. Spillantini, M. G. et al. Filamentous α-synuclein inclusions link multiple system atrophy with Parkinson's disease and dementia with Lewy bodies. *Neurosci. Lett*. **251**, 205–208 (1998).

21. Tu, P. H. et al. Glial cytoplasmic inclusions in white matter oligodendrocytes of multiple system atrophy brains contain insoluble α-synuclein. *Ann. Neurol*. **44**, 415–422 (1998).
22. Kato, S. & Nakamura, H. Cytoplasmic argyrophilic inclusions in neurons of pontine nuclei in patients with olivopontocerebellar atrophy: immunohistochemical and ultrastructural studies. *Acta Neuropathol*. **79**, 584–594 (1990).
23. Petrovic, I. N. et al. Multiple system atrophy–parkinsonism with slow progression and prolonged survival: a diagnostic catch. *Mov. Disord*. **27**, 1186–1190 (2012).
24. Davidson, W. S., Jonas, A., Clayton, D. F. & George, J. M. Stabilization of α-synuclein secondary structure upon binding to synthetic membranes. *J. Biol. Chem*. **273**, 9443–9449 (1998).
25. Uéda, K. et al. Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc. Natl Acad. Sci. USA* **90**, 11282–11286 (1993).
26. Li, H. T., Du, H. N., Tang, L., Hu, J. & Hu, H. Y. Structural transformation and aggregation of human α-synuclein in trifluoroethanol: non-amyloid component sequence is essential and β-sheet formation is prerequisite to aggregation. *Biopolymers* **64**, 221–226 (2002).
27. Crowther, R. A., Jakes, R., Spillantini, M. G. & Goedert, M. Synthetic filaments assembled from C-terminally truncated α-synuclein. *FEBS Lett*. **436**, 309–312 (1998).
28. Conway, K. A., Harper, J. D. & Lansbury, P. T. Jr. Fibrils formed in vitro from α-synuclein and two mutant forms linked to Parkinson's disease are typical amyloid. *Biochemistry* **39**, 2552–2563 (2000).
29. Serpell, L. C., Berriman, J., Jakes, R., Goedert, M. & Crowther, R. A. Fiber diffraction of synthetic α-synuclein filaments shows amyloid-like cross-β conformation. *Proc. Natl Acad. Sci. USA* **97**, 4897–4902 (2000).
30. Miake, H., Mizusawa, H., Iwatsubo, T. & Hasegawa, M. Biochemical characterization of the core structure of α-synuclein filaments. *J. Biol. Chem*. **277**, 19213–19219 (2002).
31. Mougenot, A. L. et al. Prion-like acceleration of a synucleinopathy in a transgenic mouse model. *Neurobiol. Aging* **33**, 2225–2228 (2012).
32. Luk, K. C. et al. Pathological α-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. *Science* **338**, 949–953 (2012).
33. Masuda-Suzukake, M. et al. Prion-like spreading of pathological α-synuclein in brain. *Brain* **136**, 1128–1138 (2013).
34. Osterberg, V. R. et al. Progressive aggregation of α-synuclein and selective degeneration of Lewy inclusion-bearing neurons in a mouse model of parkinsonism. *Cell Rep*. **10**, 1252–1260 (2015).
35. Peelaerts, W. et al. α-Synuclein strains cause distinct synucleinopathies after local and systemic administration. *Nature* **522**, 340–344 (2015).
36. Peng, C. et al. Cellular milieu imparts distinct pathological α-synuclein strains in α-synucleinopathies. *Nature* **557**, 558–563 (2018).
37. Prusiner, S. B. et al. Evidence for α-synuclein prions causing multiple system atrophy in humans with parkinsonism. *Proc. Natl Acad. Sci. USA* **112**, E5308–E5317 (2015).
38. Tarutani, A., Arai, T., Murayama, S., Hisanaga, S. I. & Hasegawa, M. Potent prion-like behaviors of pathogenic α-synuclein and evaluation of inactivation methods. *Acta Neuropathol. Commun*. **6**, 29 (2018).
39. Yamasaki, T. R. et al. Parkinson's disease and multiple system atrophy have distinct α-synuclein seed characteristics. *J. Biol. Chem*. **294**, 1045–1058 (2019).
40. Lavenir, I. et al. Silver staining (Campbell–Switzer) of neuronal α-synuclein assemblies induced by multiple system atrophy and Parkinson's disease brain extracts in transgenic mice. *Acta Neuropathol. Commun*. **7**, 148 (2019).
41. Klingstedt, T. et al. Luminescent conjugated oligothiophenes distinguish between α-synuclein assemblies of Parkinson's disease and multiple system atrophy. *Acta Neuropathol. Commun*. **7**, 193 (2019).
42. Strohäker, T. et al. Structural heterogeneity of α-synuclein fibrils amplified from patient brain extracts. *Nat. Commun*. **10**, 5535 (2019).
43. Shahnawaz, M. et al. Discriminating α-synuclein strains in Parkinson's disease and multiple system atrophy. *Nature* **578**, 273–277 (2020).
44. Campbell, B. C. V. et al. The solubility of α-synuclein in multiple system atrophy differs from that of dementia with Lewy bodies and Parkinson's disease. *J. Neurochem*. **76**, 87–96 (2001).
45. Fujiwara, H. et al. α-Synuclein is phosphorylated in synucleinopathy lesions. *Nat. Cell Biol*. **4**, 160–164 (2002).
46. Spillantini, M. G., Crowther, R. A., Jakes, R., Hasegawa, M. & Goedert, M. α-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies. *Proc. Natl Acad. Sci. USA* **95**, 6469–6473 (1998).
47. Tuttle, M. D. et al. Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nat. Struct. Mol. Biol*. **23**, 409–415 (2016).
48. Guerrero-Ferreira, R. et al. Cryo-EM structure of alpha-synuclein fibrils. *eLife* **7**, e36402 (2018).
49. Li, Y. et al. Amyloid fibril structure of α-synuclein determined by cryo-electron microscopy. *Cell Res*. **28**, 897–903 (2018).
50. Li, B. et al. Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun*. **9**, 3609 (2018).
51. Guerrero-Ferreira, R. et al. Two new polymorphic structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *eLife* **8**, e48907 (2019).
52. Boyer, D. R. et al. Structures of fibrils formed by α-synuclein hereditary disease mutant H50Q reveal new polymorphs. *Nat. Struct. Mol. Biol*. **26**, 1044–1052 (2019).
53. Boyer, D. R. et al. The α-synuclein hereditary mutation E46K unlocks a more stable, pathogenic fibril structure. *Proc. Natl Acad. Sci. USA* **117**, 3592–3602 (2020).
54. Zhao, K. et al. Parkinson's disease associated mutation E46K of α-synuclein triggers the formation of a novel fibril structure. Preprint at https://www.biorxiv.org/content/10.1101/870758v2.
55. Sangwan, S. et al. Inhibition of synucleinopathic seeding by rationally designed inhibitors. *eLife* **9**, e46775 (2020).
56. Goedert, M., Falcon, B., Zhang, W., Ghetti, B. & Scheres, S. H. W. Distinct conformers of assembled tau in Alzheimer's and Pick's diseases. *Cold Spring Harb. Symp. Quant. Biol*. **83**, 163–171 (2018).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Clinical history and neuropathology

MSA case 1 was in an 85-year-old woman who died with a neuropathologically confirmed diagnosis of MSA-P following a 9-year history of bradykinesia, rigidity in upper and lower limbs and autonomic failure. MSA case 2 was in a 68-year-old woman who died with a neuropathologically confirmed diagnosis of MSA-C following an 18-year history of cerebellar ataxia, gait disturbance and autonomic failure. MSA case 3 was in a 59-year-old man who died with a neuropathologically confirmed diagnosis of MSA-C following a 9-year history of dysarthria, cerebellar ataxia and autonomic failure. MSA case 4 was in a 64-year-old man who died with a neuropathologically confirmed diagnosis of MSA-C following a 10-year history of cerebellar ataxia, dysarthria and autonomic failure. MSA case 5 was in a 70-year-old man who died with a neuropathologically confirmed diagnosis of MSA-C following a 19-year history of cerebellar ataxia and autonomic failure. DLB case 1 was in a 59-year old man who died with a neuropathologically confirmed diagnosis of DLB following a 10-year history of resting tremor, bradykinesia, rigidity, postural instability and visual hallucinations. DLB case 2 was in a 74-year old man who died with a neuropathologically confirmed diagnosis of diffuse Lewy body disease following a 13-year history of bradykinesia, postural instability and visual hallucinations. DLB case 3 was in a 78-year old man who died with a neuropathologically confirmed diagnosis of diffuse Lewy body disease following a 15-year history of resting tremor, bradykinesia, autonomic symptoms and visual hallucinations.

Extraction of α-synuclein filaments

Sarkosyl-insoluble material was extracted from fresh-frozen brain regions of individuals with MSA and DLB, essentially as previously described[38]. In brief, tissues were homogenized in 20 vol (v/w) extraction buffer consisting of 10 mM Tris-HCl, pH 7.5, 0.8 M NaCl, 10% sucrose and 1 mM EGTA. Homogenates were brought to 2% sarkosyl and incubated for 30 min. at 37 °C. Following a 10 min centrifugation at 10,000*g*, the supernatants were spun at 100,000*g* for 20 min. The pellets were resuspended in 500 µl/g extraction buffer and centrifuged at 3,000*g* for 5 min. The supernatants were diluted threefold in 50 mM Tris-HCl, pH 7.5, containing 0.15 M NaCl, 10% sucrose and 0.2% sarkosyl, and spun at 166,000*g* for 30 min. Sarkosyl-insoluble pellets were resuspended in 100 µl/g of 30 mM Tris-HCl, pH 7.4. We used approximately 0.5 g tissue for cryo-EM and 0.5 g for negative-stain immuno-electron microscopy. In some experiments, sarkosyl-insoluble pellets were resuspended in 30 mM Tris-HCl, 2% SDS, left at room temperature for 30 min and spun at 100,000*g* for 30 min. The pellets were resuspended in 8 M urea. Both supernatants and pellets were immunoblotted using anti-pS129 α-synuclein antibody.

Immunolabelling and histology

Immunogold negative-stain electron microscopy and western blotting were carried out as previously described[57]. Filaments were extracted from putamen in MSA cases 1–5, the frontal cortex in MSA cases 1, 2, 3 and 5, the cerebellum in MSA case 1, the frontal cortex in DLB cases 1 and 2, and the amygdala in DLB case 3. PER4[46], a rabbit polyclonal serum that recognizes the carboxy-terminal region of α-synuclein, was used at 1:50. Images were acquired at 11,000× with a Gatan Orius SC200B CCD detector on a Tecnai G2 Spirit at 120 kV. For western blotting, the samples were resolved on 4–12% Bis-Tris gels (NuPage) and the primary antibodies diluted in PBS plus 0.1% Tween 20 and 5% non-fat dry milk. Before blocking, membranes were fixed with 1% paraformaldehyde for 30 min. Primary antibodies were: Syn303 (a mouse monoclonal antibody that recognizes the N-terminus of α-synuclein) (BioLegend)[58] at 1:4,000, PER4 at 1:4,000 and pS129 (a rabbit monoclonal antibody that recognizes α-synuclein phosphorylated at S129) (ab51253, Abcam) at 1:5,000. Histology and immunohistochemistry were carried out as previously described[38,59]. Some brain sections (8 µm) were counterstained with haematoxylin. The primary antibody was specific for α-synuclein phosphorylated at S129 (ab51253).

Seeded α-synuclein aggregation

The ability of sarkosyl-insoluble fractions from the putamen in MSA cases 1–5 to convert expressed soluble α-synuclein into its abnormal form was examined, as previously described[60,61]. Following the addition of variable amounts of seeds (ranging from 1 to 4,700 pg/ml), transfected SH-SY5Y cells were incubated for three days. Sarkosyl-insoluble α-synuclein was extracted, run on 15% SDS–PAGE and immunoblotted with a mouse monoclonal antibody specific for α-synuclein phosphorylated at S129 (pSyn64 at 1:1,000)[45]. Band intensities were quantified using ImageJ software.

Mass spectrometry of sarkosyl-insoluble α-synuclein

Protease digestion and nano-flow liquid chromatography–ion trap tandem mass spectrometry (LC-MS/MS) (Thomas Scientific, Q Exactive HF) were used to identify post-translational modifications in sarkosyl-insoluble α-synuclein extracted from the putamen in MSA cases 1–5[62]. The concentration of α-synuclein was determined using an enzyme-linked immunosorbent assay kit (Abcam). Sarkosyl-insoluble fractions containing approximately 65 ng of α-synuclein were treated with 70% formic acid for 1 h at room temperature, diluted in water and dried. They were digested overnight with trypsin and lysyl-endopeptidase. Peptides were then analysed by LC-MS/MS.

Cryo-EM

Extracted α-synuclein filaments were applied to glow-discharged holey carbon gold grids (Quantifoil R1.2/1.3, 300 mesh) and plunge-frozen in liquid ethane using an FEI Vitrobot Mark IV. Micrographs were acquired using two different Thermo Fisher Titan Krios microscopes that were operated at 300 kV. On the first microscope, at the MRC Laboratory of Molecular Biology, a Gatan K2 Summit direct detector was used in counting mode. On the second microscope, at the UK electron Bio-Imaging Centre (eBIC), a Gatan K3 direct detector in super-resolution mode was used. Inelastically scattered electrons were removed by a GIF Quantum energy filter (Gatan) using a slit width of 20 eV. Further details are given in Extended Data Table 1 and Supplementary Tables 1–3.

Helical reconstruction

Movie frames were corrected for beam-induced motion and dose-weighted using the motion-correction implementation of RELION[63]. Super-resolution K3 movies were Fourier-cropped during motion correction, and the reported pixel sizes in Extended Data Table 1 and Supplementary Tables 1–3 are the physical pixel sizes. Aligned, non-dose-weighted micrographs were used to estimate the contrast transfer function using CTFFIND-4.1[64]. All subsequent image-processing steps were performed using helical reconstruction methods in RELION 3.0[65]. Filaments were picked manually.

MSA datasets

Segments for reference-free 2D classification comprising an entire helical crossover were extracted using an inter-box distance of 14.1 Å. For samples extracted from the putamen, segments with a box size of 750 pixels and a pixel size of 1.15 Å were downscaled to 256 pixels for MSA cases 2–5, and segments with a box size of 900 pixels and a pixel size of 0.83 Å were downscaled to 300 pixels for MSA case 1. For samples extracted from frontal cortex in MSA cases 1, 2, 3 and 5, and cerebellum in MSA case 1, segments with a box size of 750 pixels and a pixel size of 1.15 Å were downscaled to 256 pixels. MSA type I and type II filaments from the putamen were initially separated by reference-free 2D classification, and segments that contributed to suboptimal 2D class averages were discarded. For both types of filaments, an initial helical twist of −1.4° was calculated from the apparent crossover distances of filaments in micrographs, and the helical rise was fixed at 4.7 Å. Using these values, initial 3D models for both types were constructed de novo from 2D class averages of segments that comprise entire helical crossovers using the relion_helix_inimodel2d program [65]. Type I and type II filament segments were then re-extracted using box sizes of 220 pixels for MSA cases 2–5 or 320 pixels for MSA case 1, without downscaling. Starting with these segments and an initial de novo model that was low-pass-filtered to 10 Å, 3D auto-refinement was carried out for several rounds with optimization of helical twist and rise after reconstructions showed separation of β-strands along the helical axis. We then performed Bayesian polishing and contrast transfer function refinement, followed by 3D classification with local optimization of helical twist and rise, but without further image alignment, to remove segments that yielded suboptimal 3D reconstructions. To further separate the subtypes of type II filaments, segments from MSA case 2 were subjected to additional supervised and focused 3D classifications of K45–V95 from PF-IIB; type II_1 and II_2 filaments served as references. For all cases, selected segments were used for further 3D auto-refinement. Final reconstructions were sharpened using the standard post-processing procedures in RELION[63]. Overall resolution estimates were calculated from Fourier shell correlations at 0.143 between the two independently refined half-maps, using phase-randomization to correct for convolution effects of a generous, soft-edged solvent mask that extended to 20% of the height of the box. Local resolution estimates were obtained using the same phase-randomization procedure, but with a soft spherical mask that was moved over the entire map. Using the relion_helix_toolbox program[66], helical symmetries were imposed on the post-processed maps, which were then used for model building. The reported ratios of MSA type I and type II filament segments in each case were determined by 2D classification of mixed sets of segments, which were re-extracted with box sizes of 750 or 900 pixels, while keeping the alignment parameters fixed to those resulting from the initial 3D refinements.

DLB datasets

DLB filament segments were extracted using an inter-box distance of 14.1 Å. For DLB cases 1–3, segments with a box size of 540 pixels and a pixel size of 1.15 Å were downscaled to 270 pixels. Reference-free 2D classification was performed using standard procedures.

Model building and refinement

Atomic models for type I and type II filaments were built de novo in Coot[67], using the maps of MSA case 1 and MSA case 2, respectively. Model building was started from the topologically conserved C-terminal bodies using the cryo-EM structure of recombinant α-synuclein filaments (RCSB Protein Data Bank code (PDB) 6A6B) as an initial reference[48]. The handedness of the final models was confirmed by the presence of densities for the main-chain carbonyl oxygen atoms in the map of type I filaments at a resolution of 2.6 Å. For turns with weaker densities, models were built at low display thresholds. Models containing five β-sheet rungs were refined in real-space by PHENIX using local symmetry and geometry restraints[68]. MolProbity[69] was used for model validation. Additional details are given in Extended Data Table 1.

Ethical review processes and informed consent

The studies carried out at Tokyo Metropolitan Institute of Medical Science and at Indiana University were approved through the ethical review processes of each Institution. Informed consent was obtained from the patients' next of kin.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw cryo-EM micrographs are available in EMPIAR, entry numbers EMPIAR-10357 (MSA case 1) and EMPIAR-10358 (MSA case 2). Cryo-EM maps have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-10650 for type I filaments from MSA case 1, EMD-10651 for type II_1 filaments from MSA case 2 and EMD-10652 for type II_2 filaments from MSA case 2. The corresponding atomic models have been deposited in the Protein Data Bank under the following accession numbers: 6XYO for type I filaments from MSA case 1, 6XYP for type II_1 filaments from MSA case 2 and 6XYQ for type II_2 filaments from MSA case 2. LC–MS/MS data were obtained from the ProteomeXchange database and have been deposited in Japan Proteome Standard Repository/Database (JPOST) under the identifier PXD018434.

57. Goedert, M., Spillantini, M. G., Cairns, N. J. & Crowther, R. A. Tau proteins of Alzheimer paired helical filaments: abnormal phosphorylation of all six brain isoforms. *Neuron* **8**, 159–168 (1992).
58. Giasson, B. I. et al. A panel of epitope-specific antibodies detects protein domains distributed throughout human α-synuclein in Lewy bodies of Parkinson's disease. *J. Neurosci. Res.* **59**, 528–533 (2000).
59. Spina, S. et al. The tauopathy associated with mutation +3 in intron 10 of *Tau*: characterization of the MSTD family. *Brain* **131**, 72–89 (2008).
60. Nonaka, T., Watanabe, S. T., Iwatsubo, T. & Hasegawa, M. Seeded aggregation and toxicity of α-synuclein and tau: cellular models of neurodegenerative diseases. *J. Biol. Chem.* **285**, 34885–34898 (2010).
61. Tarutani, A. et al. The effect of fragmented pathogenic α-synuclein seeds on prion-like propagation. *J. Biol. Chem.* **291**, 18675–18688 (2016).
62. Kametani, F. et al. Mass spectrometric analysis of accumulated TDP-43 in amyotrophic lateral sclerosis brains. *Sci. Rep.* **6**, 23281 (2016).
63. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
64. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
65. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
66. Scheres, S. H. W. Amyloid structure determination in RELION-3.1. *Acta Crystallogr. D*. **76**, 94–101 (2020).
67. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
68. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
69. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

Acknowledgements We thank the families of the patients for donating brain tissues; T. Nakane for help with RELION; T. Darling and J. Grimmett for help with high-performance computing; F. Epperson, U. Kuederli, R. Otani and R. M. Richardson for support with neuropathology; R. A. Crowther, B. Falcon, S. Lovestam, M. G. Spillantini and W. Zhang for helpful discussions. M.G. is an Honorary Professor in the Department of Clinical Neurosciences of the University of Cambridge and an Associate Member of the UK Dementia Research Institute. This work was supported by the UK Medical Research Council (MC_UP_A025_1013, to S.H.W.S., and MC_U105184291, to M.G.), Eli Lilly and Company (to M.G.), the European Union (EU/EFPIA/Innovative Medicines Initiative [2] Joint Undertaking IMPRIND, project 116060, to M.G.), the Japan Agency for Medical Research and Development (JP18ek0109391 and JP18dm020719, to M.H.), the US National Institutes of Health (P30-AG010133 and U01-NS110437, to B.G.) and the Department of Pathology and Laboratory Medicine, Indiana University School of Medicine (to B.G.). We acknowledge Y. Chaban at Diamond for access to and support from the cryo-EM facilities at the UK electron Bio-Imaging Centre (eBIC), proposal EM17434-75, funded by the Wellcome Trust, the MRC and the BBSRC, for acquisition of the MSA case 1 dataset. This study was supported by the MRC-LMB electron microscopy facility.

Author contributions A.T., B.G., T.M., T.T., T.A., K.H., S.M., M.Y. and M.H. identified patients, performed neuropathology and extracted α-synuclein filaments from MSA cases 1–5 and DLB cases 1 and 2; M.S. extracted α-synuclein filaments from DLB case 3 and conducted immunolabelling of filaments from MSA cases 1–5 and DLB cases 1–3; F.K. and M.H. carried out mass spectrometry; A.T. and M.H. did seeded aggregation; M.S. and Y.S. performed cryo-EM; Y.S., M.S. and S.H.W.S. analysed the cryo-EM data; Y.S. and A.G.M. built the atomic models; S.H.W.S. and M.G. supervised the project; all authors contributed to writing the manuscript.

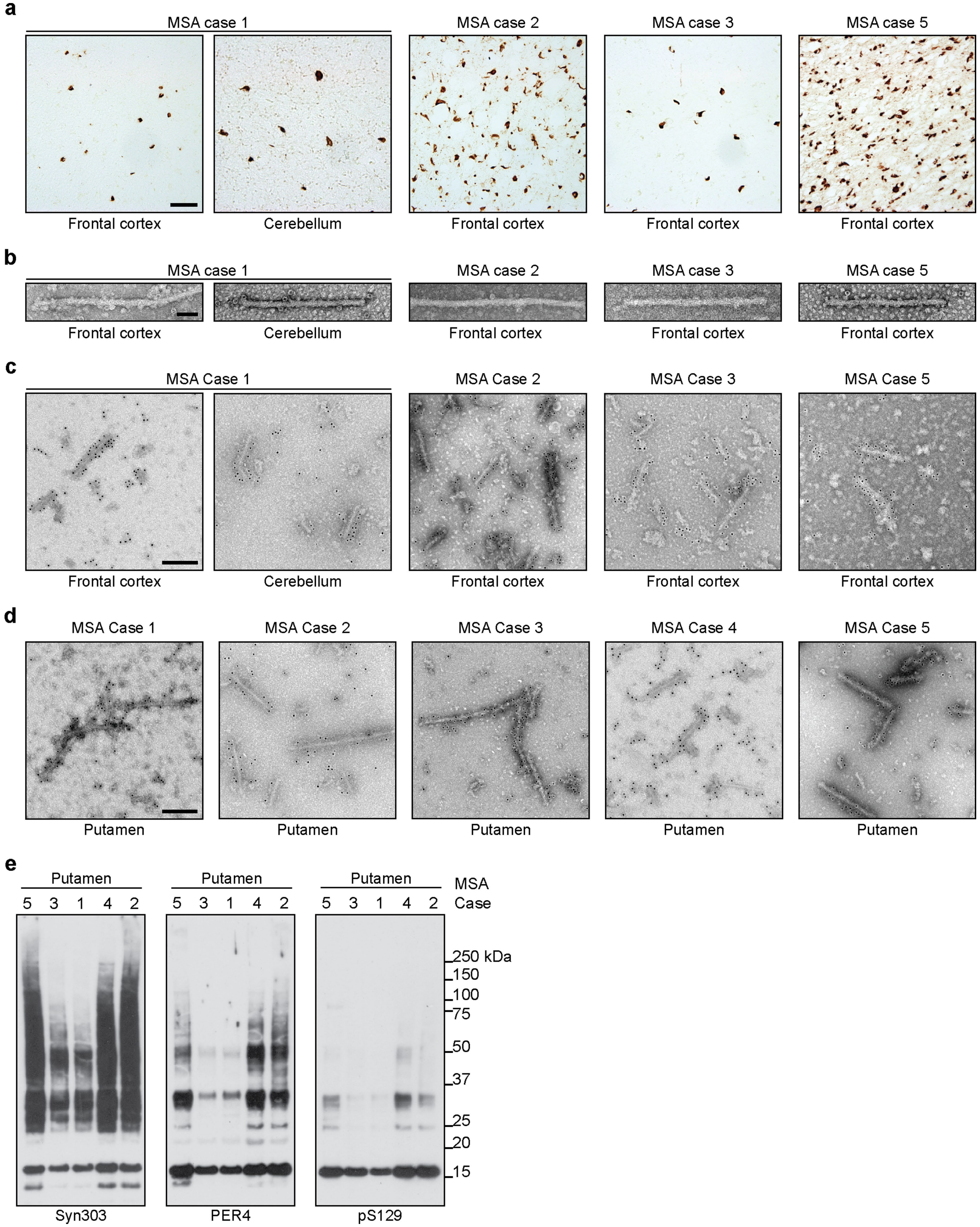
Competing interests The authors declare no competing interests.

Additional information

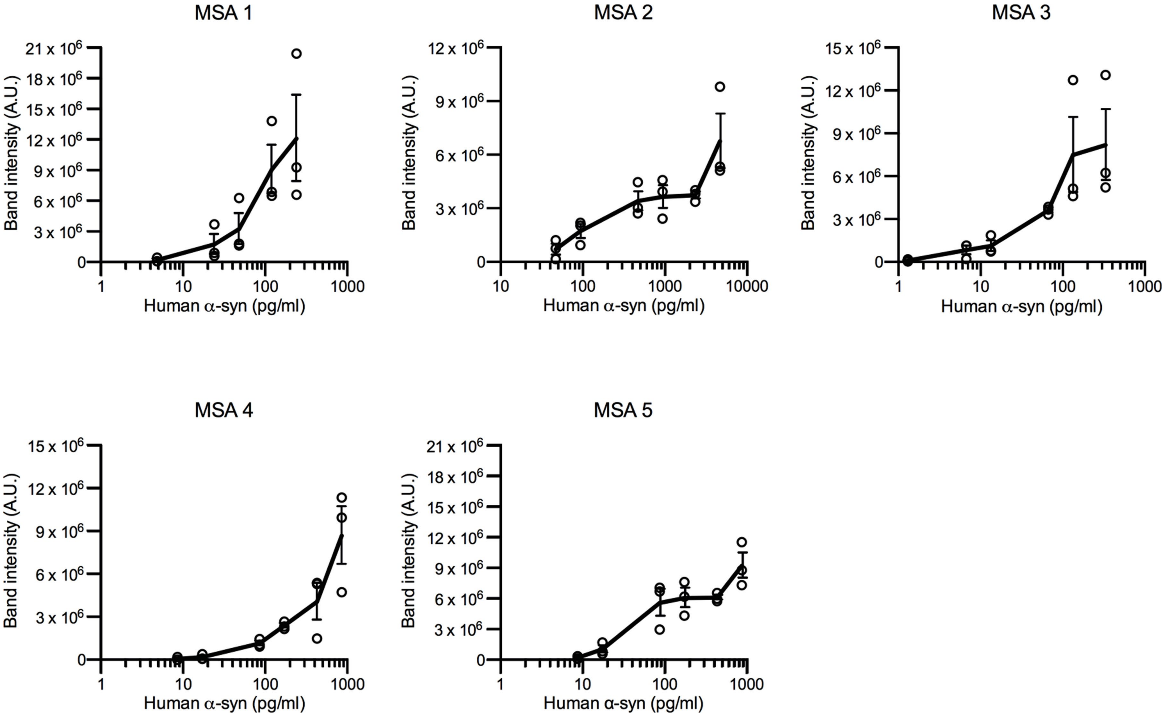
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2317-6.

Correspondence and requests for materials should be addressed to S.H.W.S. or M.G.

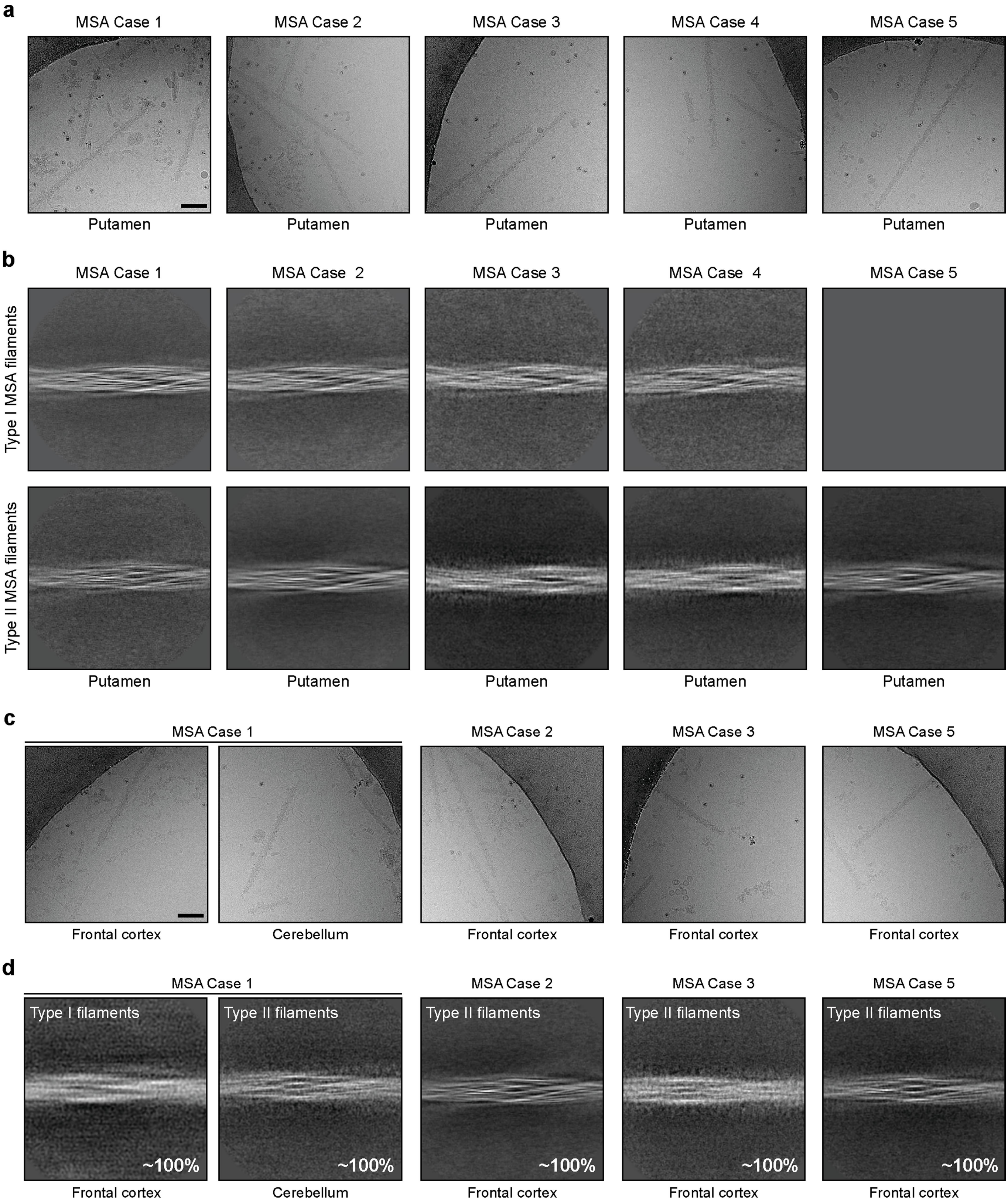
Peer review information *Nature* thanks David Eliezer, Henning Stahlberg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



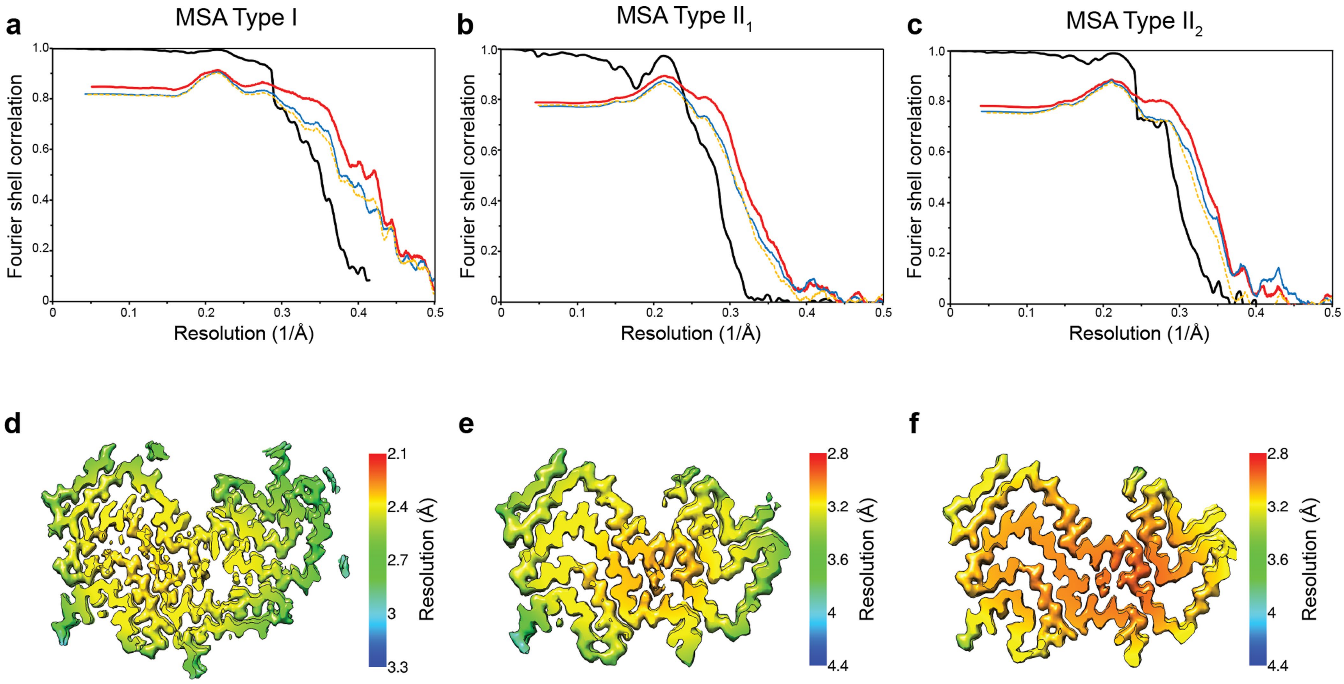
Extended Data Fig. 1 | Filamentous α-synuclein pathology and immunolabelling of α-synuclein filaments from MSA. a, Staining of inclusions in the frontal cortex in MSA cases 1, 2, 3 and 5 and the cerebellum in MSA case 1 by the antibody pS129 (brown). Scale bar, 50 μm. **b**, Negative-stain electron microscopy images of filaments from the frontal cortex in MSA cases 1, 2, 3 and 5, and the cerebellum in MSA case 1. Scale bar, 50 nm. **c**, **d**, Representative immunogold negative-stain electron microscopy images of α-synuclein filaments extracted from the frontal cortex in MSA cases 1, 2, 3 and 5, the cerebellum in MSA case 1 and the putamen in MSA cases 1–5. Filaments were labelled with the antibody PER4. Scale bar, 200 nm. **e**, Immunoblots of sarkosyl-insoluble material from the putamen for MSA cases 1–5, using the anti-α-synuclein antibodies Syn303 (N terminus), PER4 (C terminus) and pS129 (phosphorylation of S129). For gel source data, see Supplementary Fig. 1.



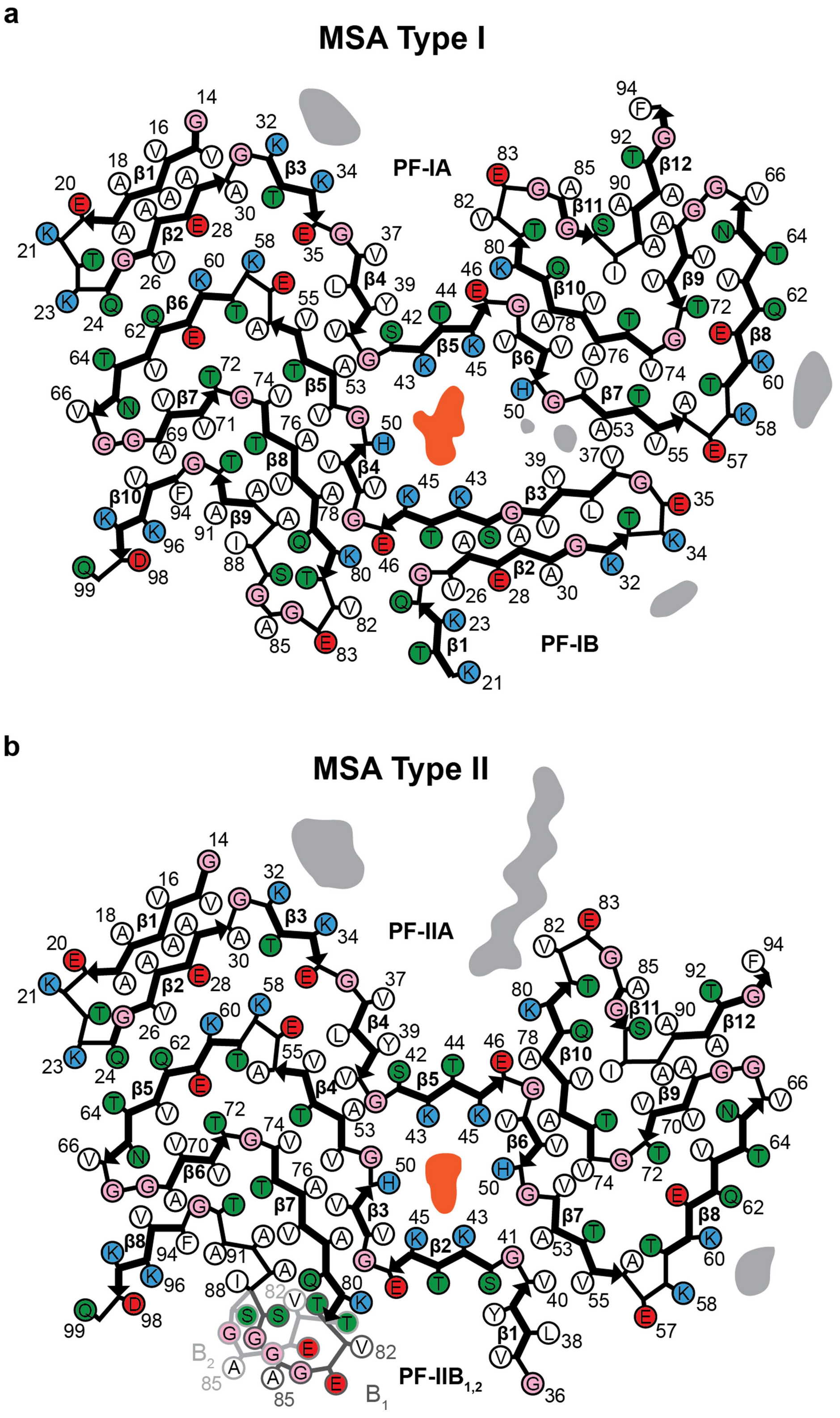
Extended Data Fig. 2 | Aggregation of α-synuclein in SH-SY5Y cells after addition of seeds from the putamen for MSA cases 1–5. Quantification of wild-type human α-synuclein phosphorylated at S129 in SH-SY5Y cells after addition of variable amounts of α-synuclein seeds from the putamen for MSA cases 1–5. The results are expressed as mean ± s.e.m. (*n* = 3 experiments).



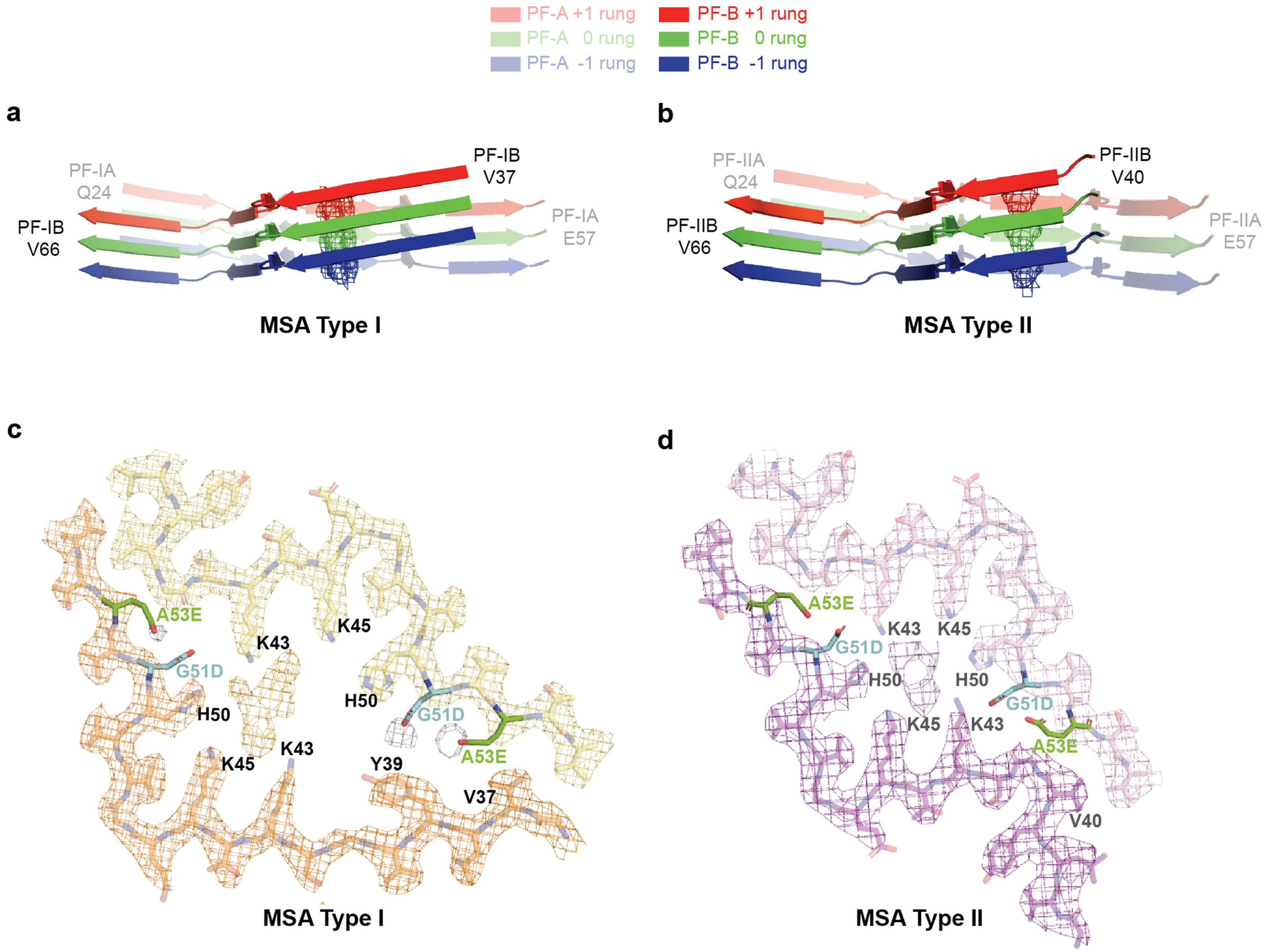
Extended Data Fig. 3 | Cryo-EM images and 2D classification of MSA filaments. **a**, **c**, Representative cryo-EM images of α-synuclein filaments from the putamen in MSA cases 1–5, the frontal cortex in MSA cases 1, 2, 3 and 5 and the cerebellum in MSA case 1. Scale bar, 100 nm. **b**, **d**, Two-dimensional class averages spanning an entire crossover of type I and type II filaments extracted from the putamen in MSA cases 1–5, the frontal cortex in MSA cases 1, 2, 3 and 5, and the cerebellum in MSA case 1.



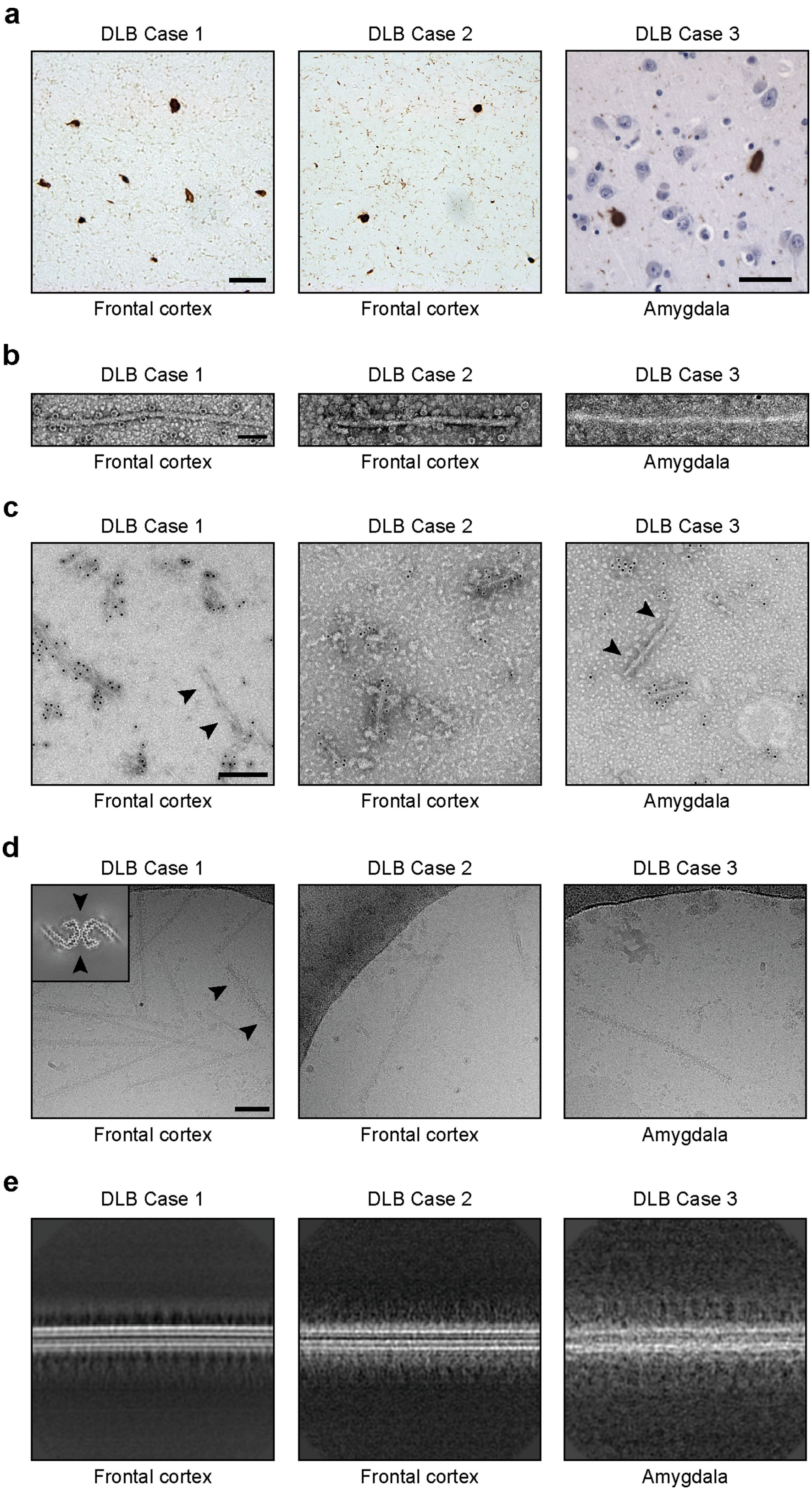
Extended Data Fig. 4 | Evaluation of the resolution of cryo-EM maps and of refined models. a–**c**, For type I (**a**), type II_1 (**b**) and type II_2 (**c**) filaments of α-synuclein from MSA: Fourier shell correlation (FSC) curves of two independently refined half-maps (black line); FSC curves of final cryo-EM reconstruction and refined atomic model (red); FSC curves of first half-map and the atomic model refined against this map (blue); FSC curves of second half-map and the atomic model refined against the first half-map (yellow dashes). **d**–**f**, Estimates of local resolution of the reconstructions of type I (**d**), type II_1 (**e**) and type II_2 (**f**) filaments of α-synuclein from MSA.



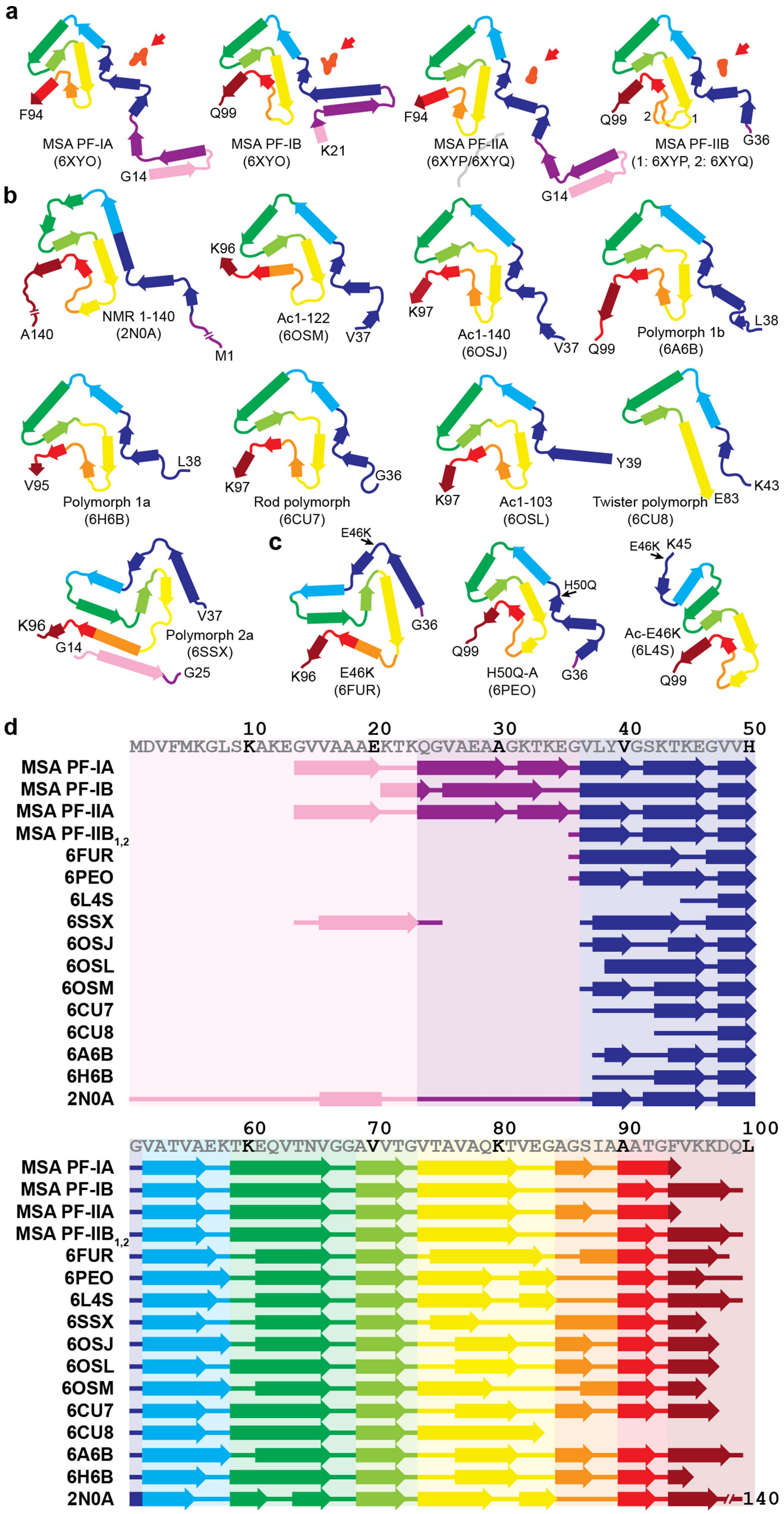
Extended Data Fig. 5 | Type I and type II filaments of α-synuclein from MSA. **a**, Schematic of a type I filament, showing asymmetric PF-IA and PF-IB. The non-proteinaceous density at the protofilament interface is shown in light red. **b**, Schematic of a type II filament, showing asymmetric PF-IIA and PF-IIB. The non-proteinaceous density at the protofilament interface is shown in light red.



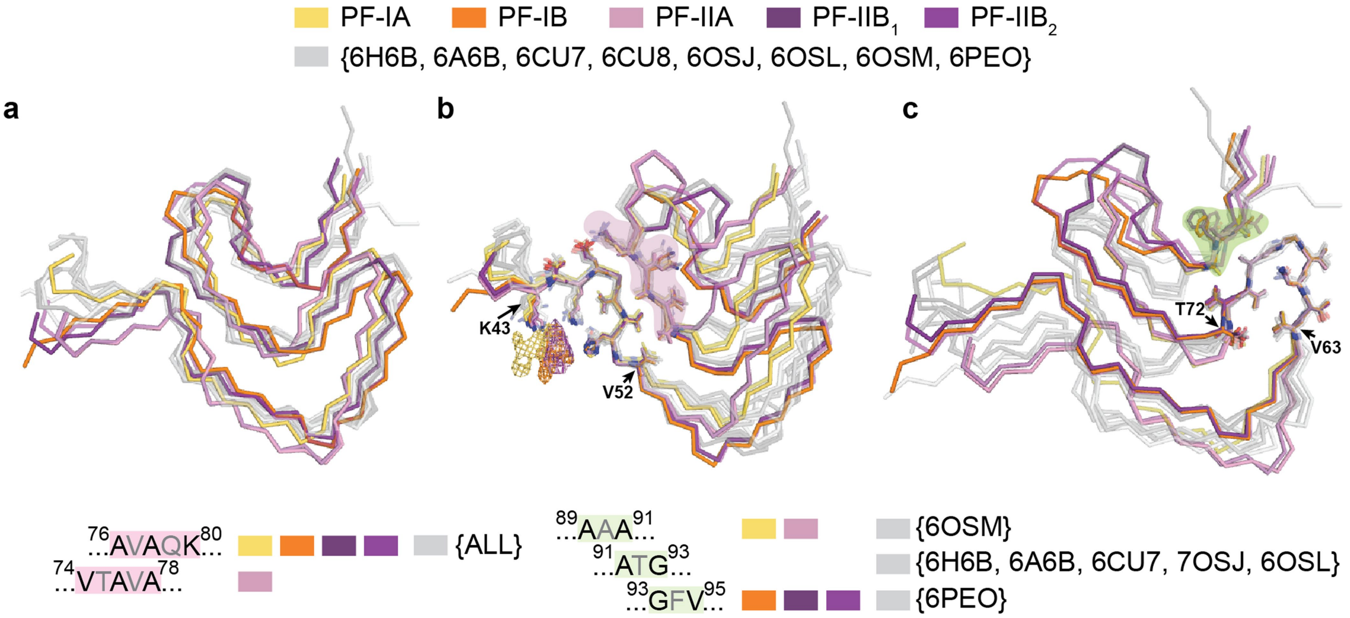
Extended Data Fig. 6 | The inter-protofilament interfaces of MSA type I and type II α-synuclein filaments. **a**, **b**, Rendered view of secondary structure elements in MSA type I (**a**) and type II (**b**) protofilament folds perpendicular to the helical axis of inter-protofilament interfaces, depicted as three rungs. Because of variations in the height of both polypeptide chains along the helical axis, each α-synuclein molecule interacts with three different molecules in the opposing protofilament. If one considers the interaction between two opposing molecules to be on the same β-sheet rung in the central cavity, the N-terminal arm of PF-IA or PF-IIA interacts with the C-terminal body of the PF-IB or PF-IIB molecule, which is one rung higher, while the C-terminal body of PF-IA or PF-IIA interacts with the N-terminal arm of the PF-IB or PF-IIB molecule, which is one rung lower. **c**, **d**, Compatibility of mutant α-synuclein (G51D and A53E) with MSA type I and type II filaments. Close-up views of atomic models of type I (**c**) and type II (**d**) α-synuclein folds containing D51 (cyan) and E53 (green). Each mutation adds two negatively charged side chains per rung in the second shell of residues around the central cavity, thus reducing the net positive charge of the shell.



Extended Data Fig. 7 | Filamentous α-synuclein pathology in DLB. a, Staining of inclusions in the frontal cortex in DLB cases 1 and 2 and the amygdala in DLB case 3 by the antibody pS129 (brown). Scale bar, 50 μm. **b**, Negative-stain electron microscopy images of filaments from the frontal cortex in DLB cases 1 and 2 and the amygdala in DLB case 3. Scale bar, 50 nm. **c**, Representative immunogold negative-stain electron microscopy images of α-synuclein filaments extracted from the frontal cortex in DLB cases 1 and 2 and the amygdala in DLB case 3. Filaments were labelled with the antibody PER4, which recognizes the C terminus of α-synuclein. Arrowheads point to an unlabelled tau paired helical filament. Scale bar, 200 nm. **d**, Representative cryo-EM images of α-synuclein filaments from the frontal cortex in DLB cases 1 and 2, and the amygdala in DLB case 3. Scale bar, 200 nm. Arrowheads point to a tau paired helical filament, as evidenced by a three-dimensional reconstruction (inset), calculated as previously described[6]. **e**, Two-dimensional class averages of α-synuclein filaments extracted from the frontal cortex in DLB cases 1 and 2 and the amygdala in DLB case 3.



Extended Data Fig. 8 | Structures of α-synuclein protofilament cores. a, Schematic of secondary structure elements in the α-synuclein protofilament cores of MSA. Red arrows point to the non-proteinaceous density (in light red) at protofilament interfaces. **b**, **c**, Secondary structure elements in the α-synuclein protofilament cores assembled from recombinant wild-type (**b**) and mutant (**c**) α-synuclein. β-Strands are shown as thick arrows. **d**, Schematic depicting the first 100 amino acids of human α-synuclein, comparing secondary structure elements in protofilament cores from MSA with those in protofilament cores assembled from recombinant α-synuclein. As observed previously for tau filaments[9], the arrangement of residues in β-strands is largely conserved among protofilament cores. This is especially the case for residues that adopt the conserved three-layered L-shaped motif, and less so for residues in the N-terminal arms.



Extended Data Fig. 9 | MSA filaments differ from those assembled with recombinant α-synuclein. a, Overlay of the three-layered L-shaped motifs of MSA α-synuclein filaments (yellow, orange, pink and purple) and filaments assembled in vitro using recombinant α-synuclein that contain a similar motif (grey). Despite topological similarities, none of the three-layered L-shaped motifs in recombinant α-synuclein protofilaments is identical to those of MSA protofilaments. The closest similarity to an in vitro structure is between PF-IIB$_2$ and PDB 6PEO[52], which differ only in the bend positions in the outer layer (between E57 and K58 for PF-IIB$_2$ and between T59 and K60 for PDB 6PEO). **b**, Overlay of MSA and recombinant α-synuclein structures on the basis of the turn at residues K43–V52, revealing a conserved interface between residues E46–V49 and V74–A78 or A76–K80 (red highlight), including the formation of a salt bridge between E46 and K80. **c**, Overlay of MSA and recombinant α-synuclein structures on the basis of the conserved turn at residues V63–T72, revealing a second conserved turn (V63–T72) and a conserved packing through tight interdigitations of small side chains between residues A69–T72 and residues on the inner layer (green highlight). In MSA PF-IA and PF-IIA filaments, as well as in PDB 6OSM[47], these residues are A89 and A91; in MSA PF-IB and PF-IIB filaments, as well as in PDB 6PEO, they are G93 and V95; in several recombinant α-synuclein structures, they are A91 and G93.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation

	Case 1 Type I (EMDB-10650) (PDB 6XYO) (EMPIAR-10357)	Case 1 Type II (EMPIAR-10357)	Case 2 Type I (EMPIAR-10358)	Case 2 Type II$_1$ (EMDB-10651) (PDB 6XYP) (EMPIAR-10358	Case 2 Type II$_2$ (EMDB-10652) (PDB 6XYQ) (EMPIAR-10358)
Data collection and processing					
Magnification	105,000	105,000	105,000	105,000	105,000
Voltage (kV)	300	300	300	300	300
Electron exposure (e–/Å^2)	49.2	49.2	47.5	47.5	47.5
Defocus range (μm)	-1.7 to -2.8	-1.7 to -2.8	-1.7 to -2.6	-1.7 to -2.6	-1.7 to -2.6
Pixel size (Å)	0.829	0.829	1.15	1.15	1.15
Symmetry imposed	C1	C1	C1	C1	C1
Initial particle images (no.)	329,477	329,477	386,301	386,301	386,301
Final particle images (no.)	120,501	34,239	10,067	23,983	93,137
Map resolution (Å)	2.60	3.68	3.61	3.29	3.09
FSC threshold	0.143	0.143	0.143	0.143	0.143
Helical twist (°)	-1.44	-1.36	-1.40	-1.41	-1.34
Helical rise (Å)	4.72	4.75	4.71	4.72	4.72
Map resolution range (Å)	2.29 to 24.12	-	-	3.05 to 28.11	2.84 to 23.00
Refinement					
Initial model used (PDB code)	6A6B	-	-	6A6B	6A6B
Model resolution (Å)	2.60	-	-	3.27	3.1
FSC threshold	0.5	-	-	0.5	0.5
Map sharpening *B* factor (Å^2)	-46.75	-	-	-62.5	-60.28
Model composition					
Non-hydrogen atoms	5,490	-	-	4,955	4,955
Protein residues	800	-	-	725	725
Ligands	-	-	-	-	-
B factors (Å^2)					
Protein	57.82	-	-	73.68	69.86
Ligand	-	-	-	-	-
R.m.s. deviations					
Bond lengths (Å)	0.008	-	-	0.008	0.008
Bond angles (°)	0.739	-	-	0.643	0.706
Validation					
MolProbity score	1.43	-	-	1.59	1.72
Clashscore	5.78	-	-	8.09	9.08
Poor rotamers (%)	0.93	-	-	0	0
Ramachandran plot					
Favored (%)	97.44	-	-	97.16	96.45
Allowed (%)	100	-	-	100	100
Disallowed (%)	0	-	-	0	0

nature research

Corresponding author(s): Sjors Scheres, Michel Goedert
Last updated by author(s): Apr 7, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	EPU v1.11.1 and v2.3.079
Data analysis	RELION v3.0, CTFFIND v4.1, COOT v0.9-pre, phenix-1.17.1-3660, MOLPROBITY v4.2, PyMOL v2.3.2, Chimera v1.8.1, ImageJ, GraphPad Prism

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw cryo-EM micrographs are available in the Elecron Microscopy Public Image Archive (EMPIAR), entry numbers EMPIAR-10357 for MSA case 1 and EMPIAR-10358 for MSA case 2. Cryo-EM maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-10650 for Type I filaments of MSA case 1, EMD-10651 and EMD-10652 for Type II1 and Type II2 filaments of MSA case 2, respectively. Refined atomic models have been deposited in the Protein Data Bank (PDB) under accession numbers 6XYO for Type I filaments for MSA case 1, 6XYP and 6XYQ for Type II1 and Type II2 filaments of MSA case 2, respectively. LC-MC/MS data have been deposited in the Japan Proteome Standard Repository/Database (JPOST) under I.D. PXD018434.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For MSA: Putamen samples from 5 cases, frontal cortex samples from 4 cases and cerebellum sample from 1 case, chosen based on availability of tissue (maximum available sample size). For DLB: Frontal cortex samples from 2 cases and amygdala sample from 1 case, chosen based on availability of tissue (maximum available sample size).
Data exclusions	Pre-established common image classification procedures (S.H.W. Scheres, J. Struc. Biol. 180: 519-530, (2012)) were employed to select particle images with the highest resolution content in the cryo-EM reconstruction process. Details of the number of selected images are given in Extended Data Table 1 and Supplementary Tables 1-3.
Replication	All attempts at replication were successful. At least three independent biological repeats per experiment where representative data is shown.
Randomization	Not relevant to study. Samples were allocated into two experimental groups (putamen, frontal cortex and cerebellum samples from cases of MSA and frontal cortex and amygdala samples from cases of DLB) based on neuropathological examination.
Blinding	Not relevant to study. Samples were allocated into two experimental groups (putamen, frontal cortex and cerebellum samples from cases of MSA and frontal cortex and amygdala samples from cases of DLB) based on neuropathological examination.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☐	☒	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used	Primary antibodies used are presented in the Methods section with validation referenced. They are: Syn303 (BioLegend MMS-5085; diluted 1:4,000 for immunoblotting), PER4 (Diluted 1:4,000 for immunoblotting and 1:50 for immunogold negative-stain EM), pS129 (Abcam ac51253; diluted 1:1,000 for immunohistochemistry, 1:5,000 for immunoblotting and 1:50 for immunogold negative-stain EM), pSyn64 (Diluted 1:1,000 for immunoblotting).
Validation	Syn303 validated against human α-synuclein N-terminus in (Giasson et al. 2000 Science 290, 985-989); PER4 validated against human α-synuclein C-terminus in (Spillantini et al. 1998 PNAS 95, 6469-6473); pS129 validated against human α-synuclein pS129 in manufacturer's datasheet (abcam); pSyn64 validated against human α-synuclein pS129 in (Fujiwara et al. 2002 Nat Cell Biol 4, 160-164).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Human neuroblastoma SH-SY5Y cell line was obtained from ATCC.
Authentication	We declare that none of the cell lines used were authenticated.

Mycoplasma contamination	The cell line used was mycoplasma-free
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics	See Methods section. Age at death: 85, 68, 59, 64, 70, 59, 74 and 78; Gender: 2x female, 6x male; Diagnoses: 5x MSA, 3x DLB.
Recruitment	Selected based on neuropathological examination. No bias was present.
Ethics oversight	The studies carried out at Tokyo Metropolitan Institute of Medical Science and at Indiana University and the University of Kansas were approved through each university's Institutional Review Board (IRB). Informed consent was obtained from the patients’ next of kin.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Author Correction: Global status and conservation potential of reef sharks

https://doi.org/10.1038/s41586-020-2692-z

Correction to: *Nature* https://doi.org/10.1038/s41586-020-2519-y

M. Aaron MacNeil, Demian D. Chapman, Michelle Heupel, Colin A. Simpfendorfer, Michael Heithaus, Mark Meekan, Euan Harvey, Jordan Goetze, Jeremy Kiszka, Mark E. Bond, Leanne M. Currey-Randall, Conrad W. Speed, C. Samantha Sherman, Matthew J. Rees, Vinay Udyawer, Kathryn I. Flowers, Gina Clementi, Jasmine Valentin-Albanese, Taylor Gorham, M. Shiham Adam, Khadeeja Ali, Fabián Pina-Amargós, Jorge A. Angulo-Valdés, Jacob Asher, Laura García Barcia, Océane Beaufort, Cecilie Benjamin, Anthony T. F. Bernard, Michael L. Berumen, Stacy Bierwagen, Erika Bonnema, Rosalind M. K. Bown, Darcy Bradley, Edd Brooks, J. Jed Brown, Dayne Buddo, Patrick Burke, Camila Cáceres, Diego Cardeñosa, Jeffrey C. Carrier, Jennifer E. Caselle, Venkatesh Charloo, Thomas Claverie, Eric Clua, Jesse E. M. Cochran, Neil Cook, Jessica Cramp, Brooke D'Alberto, Martin de Graaf, Mareike Dornhege, Andy Estep, Lanya Fanovich, Naomi F. Farabaugh, Daniel Fernando, Anna L. Flam, Camilla Floros, Virginia Fourqurean, Ricardo Garla, Kirk Gastrich, Lachlan George, Rory Graham, Tristan Guttridge, Royale S. Hardenstine, Stephen Heck, Aaron C. Henderson, Heidi Hertler, Robert Hueter, Mohini Johnson, Stacy Jupiter, Devanshi Kasana, Steven T. Kessel, Benedict Kiilu, Taratu Kirata, Baraka Kuguru, Fabian Kyne, Tim Langlois, Elodie J. I. Lédée, Steve Lindfield, Andrea Luna-Acosta, Jade Maggs, B. Mabel Manjaji-Matsumoto, Andrea Marshall, Philip Matich, Erin McCombs, Dianne McLean, Llewelyn Meggs, Stephen Moore, Sushmita Mukherji, Ryan Murray, Muslimin Kaimuddin, Stephen J. Newman, Josep Nogués, Clay Obota, Owen O'Shea, Kennedy Osuka, Yannis P. Papastamatiou, Nishan Perera, Bradley Peterson, Alessandro Ponzo, Andhika Prasetyo, L. M. Sjamsul Quamar, Jessica Quinlan, Alexei Ruiz-Abierno, Enric Sala, Melita Samoilys, Michelle Schärer-Umpierre, Audrey Schlaff, Nikola Simpson, Adam N. H. Smith, Lauren Sparks, Akshay Tanna, Rubén Torres, Michael J. Travers, Maurits van Zinnicq Bergmann, Laurent Vigliola, Juney Ward, Alexandra M. Watts, Colin Wen, Elizabeth Whitman, Aaron J. Wirsing, Aljoscha Wothke, Esteban Zarza-Gonzâlez & Joshua E. Cinner

In this Article, the first name of author 'Darcy Bradley' was misspelled as 'Darcey', and the surname of author Naomi F. Farabaugh was misspelled as 'Farabough'. The Article has been corrected online.

Publisher Correction: Structural basis of DNA targeting by a transposon-encoded CRISPR–Cas system

https://doi.org/10.1038/s41586-020-2662-5

Correction to: *Nature* https://doi.org/10.1038/s41586-019-1849-0

Tyler S. Halpin-Healy, Sanne E. Klompe, Samuel H. Sternberg & Israel S. Fernández

In this Article, Supplementary Table 1 and Supplementary Video 1 were originally not uploaded online; these files have now been uploaded to the original Article.

Publisher Correction: Two dynamically distinct circuits drive inhibition in the sensory thalamus

https://doi.org/10.1038/s41586-020-2680-3

Correction to: *Nature* https://doi.org/10.1038/s41586-020-2512-5

Rosa I. Martinez-Garcia, Bettina Voelcker, Julia B. Zaltsman, Saundra L. Patrick, Tanya R. Stevens, Barry W. Connors & Scott J. Cruikshank

In the HTML version of this Article, the affiliations of Scott J. Cruikshank should not have included ‘Present address: UAB Comprehensive Neuroscience Center, University of Alabama at Birmingham, Birmingham, AL, USA’ or ‘Present address: Center for Neural Science, New York University, New York, NY, USA’. His affiliations should have included: Department of Neuroscience at Brown University; The UAB Civitan International Research Center at University of Alabama at Birmingham; UAB Comprehensive Neuroscience Center at University of Alabama at Birmingham; and Present address: Department of Neurobiology at University of Alabama at Birmingham. The PDF and print versions were correct.

In addition, ref. 27 of the Article (Clemente-Perez et al., 2017), which was cited in relation to reduced bursting in somatostatin-Cre neurons, was not properly referenced for the description of the anatomical distribution of the cells. Ref. 27 (shown as ref. [1] to this Amendment) should have been cited for the observation that somatostatin-expressing neurons are selectively localized to the outer tiers of the thalamic reticular nucleus, and appeared after the text: “The highest densities of SOM–tdT cells were near the medial and lateral edges of the sector[1]”, and: “Cre expression in the somatosensory TRN of these mice was almost entirely restricted to the edge zones[1]”. We apologize for the omission. The original Article has been corrected online.

1. Clemente-Perez, A. et al. Publisher Correction: Distinct thalamic reticular cell types differentially modulate normal and pathological cortical rhythms. *Cell Rep*. **19**, 2130–2142 (2017).

Work

Your story Send your careers story to: naturecareerseditor@nature.com

ADAPTED FROM GETTY

GIVE YOUR BRAIN A BREAK FROM ALL THIS BUSTLE

If you can, escape the plate-spinning frenzy of online meetings by going on holiday, ideally for two weeks. **By John Tregoning.**

After months of lockdown and further months of not-quite-lockdown-but-not-back-to-normal, I've been on a holiday. And not like the fake holiday I had before (see *Nature* https://doi.org/d8hk; 2020), but a real, going-away holiday. I went to Cornwall in South West England, where, once upon a time, my ancestors mined for tin, hence my impossible-to-pronounce surname; it might also explain why I chose a career that requires you to spend long periods in the dark, figuratively at least.

Taking real time away from the laboratory has been good for me, and I feel it is essential, particularly in terms of scientific productivity. It is only by stepping back from the immediate pressures of the now and the 'must-do' list that your brain can get the space to start making deeper connections. Staying home and not working is OK, but in this new era where work and home are basically the same space, getting the separation is that much harder.

If you're reading this and thinking, "I can't take a holiday, there is too much to do," I strongly urge you to rethink. I know academia can feel like a relentless race to the bottom in terms of hours worked and days contiguously spent doing science, with influential voices shouting about their 100-hour working weeks. I know there are more tasks than can possibly be done in a day. And I know this has been worse during the pandemic when there have been even more tasks and even less time. But without a break, you can end up in a plate-spinning frenzy of online meetings and household tasks (see *Nature* **581**, 226–227; 2020).

And I'm not just advocating a long weekend here and there. It needs to be a substantial period, ideally two weeks. In my experience, one week is not enough, because by the time you have eventually wound down, you are thinking about going home again. Two weeks is enough to just do nothing, which is a balm.

This, sadly, won't be possible for everyone. I know I've been extremely fortunate in being able to get away. I was lucky enough to choose somewhere that wasn't shut down over increased COVID-19 restrictions. And not everyone can afford to get away, both in terms of time away from the lab and money.

What was a bit surprising was how different being on holiday was from being at home. Having spent nearly every day since 27 March in the same house as my family, I didn't think that spending another 14 days with the same 3 people would have felt that much different. But it is all about context – days filled with ice cream and beaches rather than home-schooling and work are considerably nicer.

In addition to the benefits of the holiday itself, one important advantage for me is bracketing periods of time. It's nice to say it's only X weeks until holiday – the anticipation relieves the monotony. This is particularly valuable now, when every day is exactly the same.

Another benefit of going away is that it changes how you feel when you get back. However, this year, re-entry to work has been a bit muted: after returning, nothing really had changed. If anything, with a possible second wave in the United Kingdom coming, and with pockets of increased restrictions, it feels like we have gone back a few steps.

Particularly concerning is the new school year. I don't think things are necessarily going to return to normal for parents any time soon. This feeling of a muted re-entry is emphasized by returning not to working at work, but to working from home. And, this year, I don't get to physically see my team right after my holiday: bragging about how nice being away has been while sharing token gifts bought in the airport is normally one way of prolonging the holiday glow! Nevertheless, it's good to be back and refreshed.

John Tregoning is a reader in respiratory infections in the Department of Infectious Disease, Imperial College London, UK. He runs a blog on academic life.

DON'T BE HARSH IN PEER REVIEW

How to reckon with comments from reviewers who use 'being critical' as a justification to be mean. **By Jeff C. Clements**

I very much enjoy being a peer reviewer. Reviewing manuscripts allows me to stay up to date on the most-current research in my field, and I feel a sense of accomplishment when helping authors to effectively disseminate their science.

However, I have been discouraged by some comments from fellow reviewers that I've seen relayed to authors. Multiple reviews, which were shared with all reviewers, were rife with unnecessary, personal comments that merely served as subjective criticisms of the authors' competencies, rather than as constructive assessment of the research. One comment went as far as implying that the authors themselves were illogical and unintelligent.

Peer review is meant to be highly critical. Many researchers, however, don't receive proper training on being effective peer reviewers (I didn't). We know that we should be critical as reviewers, but we are rarely taught to be kind and courteous. I think that, all too often, this focus on criticism rather than compassion is interpreted as a licence to be mean.

Although some journals redact ad hominem reviewer comments, many do not, and authors commonly receive them. In my field of ecology and evolution, an analysis conducted by myself and colleagues found that 10–35% of peer reviews provided to authors contain demeaning language and 43% of reviews include at least one unprofessional comment (T. G. Gerwing *et al. Res. Integr. Peer Rev.* **5**, 9; 2020). Indeed, I've endured similar comments, including this one: "What the authors have done here I would not even consider science."

These comments can slow down the publishing process. For me, it takes much longer to respond to unprofessional comments than to constructive ones, because it's rare that such feedback provides tangible suggestions to address. Therefore, authors will spend more time thinking about and crafting responses.

> **"One comment went as far as implying that the authors themselves were illogical and unintelligent."**

More important are the damaging effects that such comments can have on authors. A *Nature* survey last year revealed that bullying is a potentially significant source of poor mental health in PhD students (see *Nature* **575**, 257–258; 2019). Personally, harsh reviewer comments have made me feel anxious and like an impostor.

When I receive harsh comments from reviewers as an author, I initially feel annoyed and slighted, so I try not to respond right away. Instead, I take some time to digest the comments and not take them personally, which allows me to respond in a more neutral tone.

What can I do if I see or receive unprofessional comments?

Sometimes, it's hard to get past the personal nature of these remarks. I then contact the relevant editors directly (some journals have policies for these instances; others do not). I do this as a reviewer if I see such comments relayed to authors, because many authors might not be comfortable doing so themselves. In my experience, editors are usually receptive to such feedback and often pass it along to the other reviewers. More authors and reviewers bringing comments that are just plain mean to the attention of editors might start changing the culture. I have provided a template for such communications on Twitter, which anyone can use (see go.nature.com/35j5kyz).

This year, I reviewed for a journal that included a 'positive comments' section, where reviewers can praise aspects of a manuscript. I try to do this wherever possible in my own reviews, but journals having this section as part of the review structure will help reviewers to provide uplifting comments.

When I work as a co-editor for scientific publications at Fisheries and Oceans Canada in Ottawa, where I also work as a research scientist, I do not edit original reviewer text. Instead, I send unprofessional reviews back for revision and specifically point out problems in a non-judgemental way. Having more authors and reviewers bring such issues directly to the attention of editors can, I think, facilitate more editors to do this.

Some journals are experimenting with publishing the full text of peer reviews in a manuscript. This could help to raise awareness of the problem, but because reviewers' identities are hidden, there might still be little reason for them to be courteous.

Alongside the personal steps that individual reviewers can take, proper instruction and training on how to review manuscripts constructively, collegially and courteously would go a long way. Such training could be integrated into 'research methods'-type courses in graduate school or offered as institutional workshops. I did a course on writing a good paper; why not a course on how to peer review?

In this dark and strange global pandemic, there is perhaps no better time to actively promote and foster the power of compassion in peer review – not just for the sake of science, but for the people who do it.

Jeff Clements is a government scientist at Fisheries and Oceans Canada, in Moncton.
e-mail: jefferycclements@gmail.com

that spending another 14 days with the same 3 people would have felt that much different. But it is all about context – days filled with ice cream and beaches rather than home-schooling and work are considerably nicer.

In addition to the benefits of the holiday itself, one important advantage for me is bracketing periods of time. It's nice to say it's only X weeks until holiday – the anticipation relieves the monotony. This is particularly valuable now, when every day is exactly the same.

Another benefit of going away is that it changes how you feel when you get back. However, this year, re-entry to work has been a bit muted: after returning, nothing really had changed. If anything, with a possible second wave in the United Kingdom coming, and with pockets of increased restrictions, it feels like we have gone back a few steps.

Particularly concerning is the new school year. I don't think things are necessarily going to return to normal for parents any time soon. This feeling of a muted re-entry is emphasized by returning not to working at work, but to working from home. And, this year, I don't get to physically see my team right after my holiday: bragging about how nice being away has been while sharing token gifts bought in the airport is normally one way of prolonging the holiday glow! Nevertheless, it's good to be back and refreshed.

John Tregoning is a reader in respiratory infections in the Department of Infectious Disease, Imperial College London, UK. He runs a blog on academic life.

DON'T BE HARSH IN PEER REVIEW

How to reckon with comments from reviewers who use 'being critical' as a justification to be mean. **By Jeff C. Clements**

I very much enjoy being a peer reviewer. Reviewing manuscripts allows me to stay up to date on the most-current research in my field, and I feel a sense of accomplishment when helping authors to effectively disseminate their science.

However, I have been discouraged by some comments from fellow reviewers that I've seen relayed to authors. Multiple reviews, which were shared with all reviewers, were rife with unnecessary, personal comments that merely served as subjective criticisms of the authors' competencies, rather than as constructive assessment of the research. One comment went as far as implying that the authors themselves were illogical and unintelligent.

Peer review is meant to be highly critical. Many researchers, however, don't receive proper training on being effective peer reviewers (I didn't). We know that we should be critical as reviewers, but we are rarely taught to be kind and courteous. I think that, all too often, this focus on criticism rather than compassion is interpreted as a licence to be mean.

Although some journals redact ad hominem reviewer comments, many do not, and authors commonly receive them. In my field of ecology and evolution, an analysis conducted by myself and colleagues found that 10–35% of peer reviews provided to authors contain demeaning language and 43% of reviews include at least one unprofessional comment (T. G. Gerwing *et al. Res. Integr. Peer Rev.* **5**, 9; 2020). Indeed, I've endured similar comments, including this one: "What the authors have done here I would not even consider science."

These comments can slow down the publishing process. For me, it takes much longer to respond to unprofessional comments than to constructive ones, because it's rare that such feedback provides tangible suggestions to address. Therefore, authors will spend more time thinking about and crafting responses.

> **"One comment went as far as implying that the authors themselves were illogical and unintelligent."**

More important are the damaging effects that such comments can have on authors. A *Nature* survey last year revealed that bullying is a potentially significant source of poor mental health in PhD students (see *Nature* **575**, 257–258; 2019). Personally, harsh reviewer comments have made me feel anxious and like an impostor.

When I receive harsh comments from reviewers as an author, I initially feel annoyed and slighted, so I try not to respond right away. Instead, I take some time to digest the comments and not take them personally, which allows me to respond in a more neutral tone.

What can I do if I see or receive unprofessional comments?

Sometimes, it's hard to get past the personal nature of these remarks. I then contact the relevant editors directly (some journals have policies for these instances; others do not). I do this as a reviewer if I see such comments relayed to authors, because many authors might not be comfortable doing so themselves. In my experience, editors are usually receptive to such feedback and often pass it along to the other reviewers. More authors and reviewers bringing comments that are just plain mean to the attention of editors might start changing the culture. I have provided a template for such communications on Twitter, which anyone can use (see go.nature.com/35j5kyz).

This year, I reviewed for a journal that included a 'positive comments' section, where reviewers can praise aspects of a manuscript. I try to do this wherever possible in my own reviews, but journals having this section as part of the review structure will help reviewers to provide uplifting comments.

When I work as a co-editor for scientific publications at Fisheries and Oceans Canada in Ottawa, where I also work as a research scientist, I do not edit original reviewer text. Instead, I send unprofessional reviews back for revision and specifically point out problems in a non-judgemental way. Having more authors and reviewers bring such issues directly to the attention of editors can, I think, facilitate more editors to do this.

Some journals are experimenting with publishing the full text of peer reviews in a manuscript. This could help to raise awareness of the problem, but because reviewers' identities are hidden, there might still be little reason for them to be courteous.

Alongside the personal steps that individual reviewers can take, proper instruction and training on how to review manuscripts constructively, collegially and courteously would go a long way. Such training could be integrated into 'research methods'-type courses in graduate school or offered as institutional workshops. I did a course on writing a good paper; why not a course on how to peer review?

In this dark and strange global pandemic, there is perhaps no better time to actively promote and foster the power of compassion in peer review – not just for the sake of science, but for the people who do it.

Jeff Clements is a government scientist at Fisheries and Oceans Canada, in Moncton.
e-mail: jefferycclements@gmail.com

Where I work
Jessica Frantz

It might look like someone made it in their garage, but this instrument is a really powerful directional antenna. Fellow students and I at Phoenix College in Arizona designed and built it for ASCEND, a NASA programme that funds science-education projects. In ASCEND, student teams from across the state build scientific instruments to attach to a high-altitude balloon.

Our team's idea was to live-stream the balloon's flight. To transmit video from the balloon – which can rise more than 30 kilometres into the air – we chose a 5-gigahertz radio, the kind used to supply Wi-Fi to a hotel. To get such a small, lightweight radio to transmit over long distances, we needed a really strong antenna to pick up the signal. The steel dish is so heavy that we had to attach dumbbell weights as a counterbalance.

In this photograph, taken on the morning of the launch in November 2019, we are trying to get the ground system and instruments talking to one another. But we had trouble and didn't get our live stream. It was nerve-racking, but it was a useful learning experience.

ASCEND is great because it is so hands-on – and it encourages women to speak up, which builds confidence. In August, I started at Arizona State University for the final two years of my undergraduate degree in mechanical engineering, and have joined the university's ASCEND team. Our project uses near-infrared and visible-light cameras to evaluate the health of the vegetation in Arizona's deserts.

I am quite interested in working for NASA, especially because of the Artemis mission to send people to the Moon again by 2024. And with public–private partnerships such as the SpaceX *Crew Dragon*, which carried astronauts to the International Space Station in May, there are tons of opportunities to work in aerospace and still be a part of a mission.

Jessica Frantz is an undergraduate student in mechanical engineering at Arizona State University in Tempe. **Interview by James Mitchell Crow.**

Photographed by Arsh Nadkarni.

Matters arising

Slow compression of crystalline ice at low temperature

https://doi.org/10.1038/s41586-020-2697-7

Received: 28 October 2019

Accepted: 10 August 2020

R. Bauer[1], J. S. Tse[1]✉, K. Komatsu[2], S. Machida[3] & T. Hattori[4]

ARISING FROM C. A. Tulk et al. *Nature* https://doi.org/10.1038/s41586-019-1204-5 (2019)

The first-order-like second transformation from pressure-amorphized high-density amorphous (HDA) ice[1] to the low-density amorphous (LDA) form[2] has motivated the hypothesis that a phase boundary may exist between the two distinct metastable forms and that an extension of the boundary into the supercooled water region will terminate at a critical point[3]. Thus, ordinary water consists of two liquids with very different densities[4]. A recent neutron diffracton experiment[5], however, reported a series of crystal-to-crystal transformations: when proton-disordered ice Ih was compressed at 100 K with long wait time between pressure increments, it transformed to ice IX′, then to ice XV′and then to proton-ordered ice VIII′. Remarkably, this is the same transformation sequence through the thermodynamically stable dense crystalline ices that is observed at high temperature.

We repeated the experiment[5] following a similar compression strategy but with greatly increased wait time. The results are different from those of ref. [5] but similar to two previous studies[6,7] (with some distinctions). These results help to establish an atomistic description of the amorphization process and the amorphous state structure.

We measured high-pressure, low-temperature neutron diffraction patterns using high-purity deuterated water. We performed a series of warm/cool and compression/decompression cycles to ensure no other form of ice (except ice Ih) was present. The initial pressure and temperature were 0.04 GPa and 96.8 K, respectively. The hydraulic pressure was increased in steps of 0.1 MPa min^{-1} with increments of 1 MPa, which corresponds to a ramp rate of less than 0.002 GPa min^{-1} (Supplementary Fig. S1). At each pressure point, the sample was allowed to equilibrate for 30 min before the diffraction pattern was measured. The procedure was repeated for 13 hours until the hydraulic pressure reached 20 MPa. From the equation of state of lead, the pressure on the sample was calculated to be 1 GPa. All diffraction patterns reveal only pure ice Ih. The pressure was then maintained at 1 GPa for 12 hours. During this period, the intensities of the Bragg peaks of ice Ih decreased slowly. Concomitantly, the baseline of the diffraction pattern increased, suggesting a gradual transformation to HDA ice. The pressure was held at 1 GPa and diffraction patterns were measured intermittently for 14 hours. Even after extensive annealing, there was still 15–20% of untransformed crystalline ice. The pressure was then increased in steps of 0.1 GPa. The amount of HDA increased and at 1.5 GPa, ice VII was observed, with the total disappearance of ice Ih. (We note that HDA always coexists with crystalline ice Ih or ice VII.)

We performed a second experiment using Fluorinet as a pressure-transmitting medium[8]. A comparison of the diffraction patterns (Supplementary Fig. S2) measured with and without Flourinet shows that the counting statistic and width of the Bragg peaks greatly improved without Flourinet. No crystalline contamination was found. The ice sample was again compressed in small increments with long wait time up to 2.5 GPa and then the pressure was reduced. The sequence of structural transformations and the associated pressures are identical to that of pure ice Ih. The structural changes revealed by the diffraction patterns at selected pressures are summarized in Fig. 1.

The salient structural change is succinctly summarized as follows. The sequential crystal-to-crystal transformations reported in ref. [5] were not reproduced. Ice Ih persists up to about 1 GPa. Upon prolonged relaxation, it slowly transformed to HDA ice. The amorphization transformation was not completed within one day. Upon further compression, there is a persistent presence of a small amount of residual ice Ih. At 1.5 GPa, the diffraction pattern of ice Ih disappeared and ice VII started to emerge. Above this pressure, the amount of ice VII increases at the expense of decreasing HDA. Our examination of the evolution of the three ice Ih Bragg diffractions—that is, 100, 002 and 101 within the pressure range 1.1–1.5 GPa—reveals an interesting behaviour in the peak positions and widths (Fig. 2). The 100 and 002 Bragg peaks shift to smaller *d* spacing and then increase when close to amorphization. The reverse trend is observed for the 101 reflection. Despite the reduction in intensities, the peak widths for 100 and 101 appear to be fairly

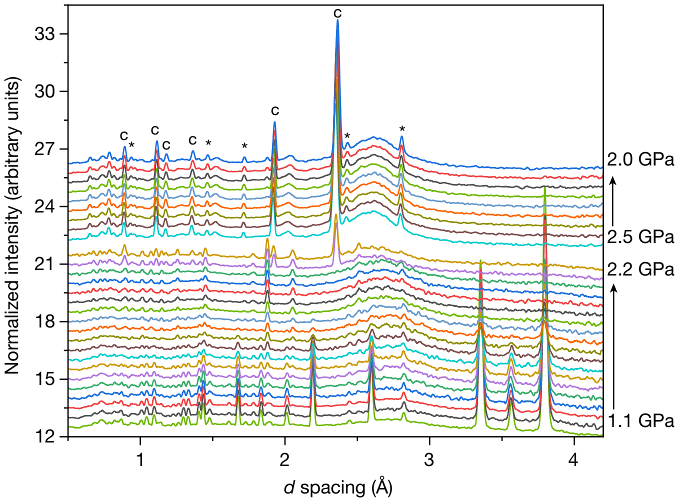


Fig. 1 | Neutron diffraction patterns of compressed ice. The lower part of the figure depicts the patterns measured from 1.1 GPa to 2.2 GPa, showing the gradual transformation from ice Ih to amorphous ice and the emergence of ice VII. The upper part of the figure shows the persistence of ice VII upon relaxation from 2.5 GPa to 2.0 GPa. The asterisks indicate the Bragg peaks of the lead calibrant. The labels 'c' highlight the crystalline Bragg peak from ice VII.

---

[1]Department of Physics and Engineering Physics, University of Saskatchewan, Saskatchewan, Saskatoon, Canada. [2]Geochemical Research Center (GCRC), Graduate School of Science, The University of Tokyo, Tokyo, Japan. [3]Neutron Science and Technology Center, CROSS, Tokai, Japan. [4]J-PARC Center, Japan Atomic Energy Agency, Naka-gun, Japan. ✉e-mail: john.tse@usask.ca

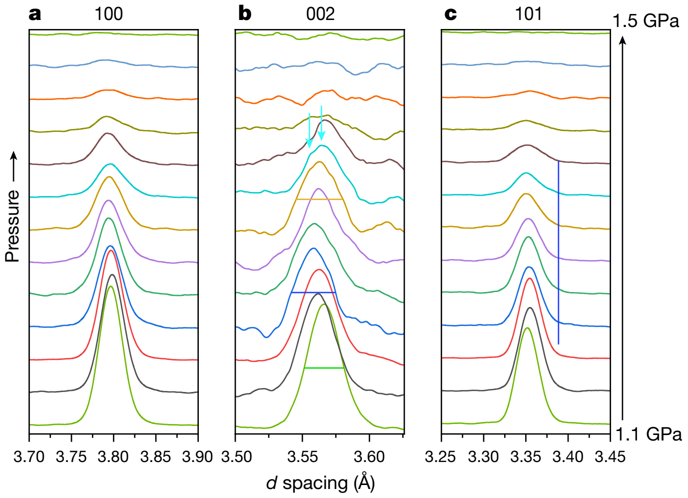


Fig. 2 | Evolution of ice Ih Bragg peak features within the pressure region from 1.1 GPa to 1.5 GPa. The evolution of the peak position and width of the 100, 002 and 101 Bragg peaks, with horizontal bars indicating the FWHM of the corresponding peak. For the 100 peak the FWHM varied between 0.028 Å and 0.03 Å; for the 101 peak, the FWHM varied from 0.026 Å to 0.029 Å. In comparison, the FWHM of the 002 peak broadened from 0.030 Å to 0.038 Å and a shoulder started to develop (indicated by arrows) at a smaller *d* spacing at pressures close to amorphization.

constant (Fig. 2a, c) with the full width at half maximum (FWHM) varying from 0.026 Å to 0.030 Å. In comparison, the width of (002) broadened from 0.030 Å to 0.038 Å and a shoulder appears just before amorphization. (Fig 2b). This observation is consistent with a previous X-ray diffraction study under quasi-hydrostatic conditions[9].

It is now possible to construct an atomistic description of the amorphization process. When ice Ih is compressed, it is metastable up to 1 GPa. The hexagonal lattice starts to distort owing to a shear instability, as predicted earlier[10] and confirmed[11] by neutron inelastic scattering experiments. This is immediately followed by the transformation to HDA. Inspection of the ice phase diagram (for example, figure 1b of ref. [5]) shows at temperatures exceeding 120 K, ice VI with interpenetrating hydrogen (H)-bond networks is the stable phase at pressures higher than 0.8 GPa. It is reasonable to suggest that the amorphization at 1 GPa and 100 K is the onset of the transition from a single H-bond network to a higher-density interpenetrating H-bond network[9]. The Raman spectrum, which probes the local structure of HDA ice, shares a strong resemblance to ice VI (ref. [12]). Analysis of the theoretical HDA structure obtained from first-principles molecular dynamics calculations confirms this suggestion[13]. At low temperatures, there is insufficient energy to overcome the activation barrier for the transformation to the crystalline structure. Thus, the transformed ice structure is frustrated, losing long-range order. Upon further compression, there is enough pressure–volume mechanical work to overcome this energy barrier, forming ice VII at 1.5 GPa. This pressure, incidentally, is close to the threshold of the region of stably coexisting ice VII/ice VIII (Fig. 1b, ref. [5]). The occurrence of proton-disordered ice VII instead of the stable thermodynamic proton-ordered ice VIII is important, indicating that the H-bond networks do not have enough time to equilibrate into the thermodynamically stable disordered form. HDA is a metastable phase and the consequence of the kinetics is related to the rearrangement of the atoms. A neutron study[7] has shown that amorphization is sustained up to 130 K. The conversion of HDA to ice VII had also been reported at about 3 GPa (ref. [6]) and 100 K. The much lower pressure observed here is due to the long wait time between pressure increments, allowing better equilibration. In the same study at 175 K (ref. [6]), ice Ih was found to transform into ices IV, V and XII at 0.4–0.7 GPa and into a mixture of ices VI and VII at 1–1.2 GPa, a sequence similar to that observed in ref. [5]. These observations are collobarated by time-resolved X-ray diffraction experiments between 100 K and 160 K, showing that the structure observed is sensitive to the temperature[14]. For example, around 150 K, if the compression rate is slow, ice Ih transforms to ice II or ice IX. On the other hand, if the compression rate is high, ice Ih transforms to HDA. It is reasonable to speculate that given an even longer relaxation time the amorphous phase may eventually be bypassed. The discrepancy between the present results and the recent report[5] is thus perhaps due to differences in the temperature control[15].

Methods

See the Supplementary Information of this Comment for more details of the Methods. High-purity distilled deuterated water (Fujifilm Wako Pure Chemical Corporation) was loaded into a null scattering TiZr cup, together with a lead calibrant. The cup is then placed in an aluminum retaining ring with the entire assembly placed in the low-temperature Mito System. Temperature control was performed by a set of two Pt thermometers, inserted in the copper rings attached to the support ring of the anvils[15], to an accuracy of ±0.5 K. Neutron diffraction patterns were measured at 26 Hz in single-frame mode at beamline BL11 (PLANET) of the Materials and Life Science Experimental Facility, Japan Proton Accelerator Research Complex.

Data availability

The data that support the findings shown in the figures are available from the corresponding author upon reasonable request.

1. Mishima, O., Calvert, L. D. & Whalley, E. 'Melting ice' I at 77 K and 10 kbar: new method of making amorphous solids. *Nature* **310**, 393–395 (1984).
2. Mishima, O., Calvert, L. D. & Whalley, E. An apparently 1st-order transition between two amorphous phases of ice induced by pressure. *Nature* **314**, 76–78 (1985).
3. Mishima, O. & Stanley, E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1998).
4. Gallo, P. et al. Water: a tale of two liquids. *Chem. Rev.* **116**, 7463–7500 (2016).
5. Tulk, C. A. et al. Absence of amorphous forms when ice is compressed at low temperature. *Nature* **569**, 542–545 (2019).
6. Klotz, S. et al. Recrystallization of HDA ice under pressure by in-situ neutron diffraction to 3.9 GPa. *Z. Kristall.* **218**, 117–122 (2003).
7. Strässle, T. et al. Temperature dependence of the pressure-induced amorphization of ice studied by high-pressure neutron diffraction to 30 K. *Phys. Rev. B* **82**, 094103 (2010).
8. Varga, T. & Wilkinson, A. P. Fluorinert as a pressure-transmitting medium for high-pressure diffraction studies. *Rev. Sci. Instrum.* **74**, 4564–4566 (2003).
9. Lin, C. et al. Kinetically controlled two-step amorphization and amorphous-amorphous transition in ice. *Phys. Rev. Lett.* **119**, 135701 (2017).
10. Tse, J. S. Mechanical instability in ice Ih. A mechanism for pressure-induced amorphization. *J. Chem. Phys.* **96**, 5482–5487 (1992).
11. Strässle, T., Saitta, A. M., Klotz, S. & Braden, M. Phonon dispersion of ice under pressure. *Phys. Rev. Lett.* **93**, 225901 (2004).
12. Loerting, T., Salzmann, C., Kohl, I., Mayer, E. & Hallbrucker, A. A second distinct structural "state" of high-density amorphous ice at 77 K and 1 bar. *Phys. Chem. Chem. Phys.* **3**, 5355–5357 (2001).
13. Yong, X., Tse, J. S. & English, N. J. optPBE-vdW density functional theory study of liquid water and pressure-induced structural evolution in ice Ih. *Can. J. Chem.* **95**, 1205–1211 (2017).
14. Lin, C. L. et al. Venture into water's no man's land: structural transformations of solid H_2O under rapid compression and decompression. *Phys. Rev. Lett.* **121**, 225703 (2018).
15. Komatsu, K. et al. Development of a new *P*–*T* controlling system for neutron-scattering experiments. *High Press. Res.* **33**, 208–213 (2013).

Author contributions J.S.T. designed the study. R.B. and J.S.T. performed the experiment (with assistance from K.K., S.M. and T.H.). R.B. and J.S.T. analysed the data. All authors were involved in the writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2697-7.
Correspondence and requests for materials should be addressed to J.S.T.
Reprints and permissions information is available at http://www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.